

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:30:39 ; Search time 4075.18 Seconds
(without alignments)
439.418 Million cell updates/sec

Title: US-09-300-482-14
Perfect score: 410

Sequence: 1 ccacgcgtccgcggtcatg.....gcaacggccgcctcttgatg 410

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba1:

2: gb_ba2:

3: gb_om:

4: gb_ov:

5: gb_pat:

6: gb_ph:

7: gb_pl1:

8: gb_pl2:

9: gb_pl3:

10: gb_pr2:

11: gb_pr3:

12: em_fun:

13: em_hum1:

14: em_hum2:

15: em_in:

16: em_om:

17: em_or:

18: em_ov:

19: em_pat:

20: em_ph:

21: em_pl:

22: em_ro:

23: em_sts:

24: em_sy:

25: em_un:

26: em_vi:

27: gb_htg1:

28: gb_htg2:

29: gb_inl:

30: gb_in2:

31: em_ba1:

32: em_ba2:

33: em_hum3:

34: em_hum4:

35: gb_pr4:

36: gb_htg3:

37: gb_htg4:

38: gb_htg5:

39: gb_htg6:

40: gb_htg7:

41: em_htg1:

42: em_htg2:

43: em_htg3:

44: em_hum5:
45: gb_pl3:
46: gb_pr5:
47: gb_htg8:
48: gb_htg9:
49: gb_htg10:
50: gb_htg11:
51: gb_htg12:
52: gb_htg13:
53: gb_htg14:
54: gb_in3:
55: gb_htg15:
56: gb_htg16:
57: gb_htg17:
58: em_htg4:
59: em_htg5:
60: em_htg6:
61: em_htg7:
62: em_hum6:
63: gb_htg18:
64: gb_htg19:
65: gb_ba3:
66: em_htg8:
67: em_htg9:
68: em_htg10:
69: em_htg11:
70: em_htg12:
71: em_htg13:
72: em_htg14:
73: em_htg15:
74: em_htg16:
75: em_htg17:
76: em_htg18:
77: em_htg19:
78: em_htg20:
79: em_htg21:
80: em_htg22:
81: em_htg23:
82: gb_pr6:
83: gb_pr7:
84: gb_htg20:
85: gb_htg21:
86: gb_htg22:
87: gb_htg23:
88: gb_ro:
89: gb_sts1:
90: gb_sts2:
91: gb_sy:
92: gb_un:
93: gb_vil:
94: gb_vil2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	394.8	96.3	1686	7	AF061838
2	389.4	95.0	634	7	AF037030
3	326	79.5	150120	8	AF001552
4	323.8	79.0	2392	7	AF037037
5	320.6	78.2	1818	7	AF061837
6	319	77.8	645	7	AF037029
7	262	63.9	96240	56	AC068900
8	230	56.1	1667	7	AB007907
9	216.2	52.7	1618	45	MSU18239
10	195.2	47.6	3158	8	LDI130772
11	183	44.6	111222	7	AC007764
12	175	42.7	79837	7	AB005233
c					

AF061838 Zea mays

AF037030 Zea mays

AF001552 Oryza sat

AF037037 Zea mays

AF061837 Zea mays

AF037029 Zea mays

AC068900 Arabidops

AB007907 Glycine m

U18239 Medicago sa

AD130772 Laminaria

AC007764 Genomic s

AB005233 Arabidops

Sat Nov 4 18:11:08 2000

13	114.8	28.0	38587	29	AC011913	AC011913 Leishmani
14	110.4	26.9	42000	57	AC073566	AC073566 Leishmani
15	109.4	26.7	1683	8	D89161	D89161 Schizosacch
16	106.8	26.0	2156	54	TBGNNG	TBGNNG T. brucei gn
17	99.2	24.2	1916	5	E13660	E13660 gDNA encodi
18	96	23.4	23773	1	AE002301	AE002301 Chlamydia
19	94.4	23.0	1953	3	OAGPDGH	OAGPDGH Ovine 6-pgd
20	93	22.7	780	8	CNS01AT6	CNS01AT6 Botrytis
21	92.8	22.6	2853	45	SC6PGDEHY	SC6PGDEHY S. cerevisia
22	92.8	22.6	2853	45	SCU17155	SCU17155 Saccharomyc
23	92.8	22.6	10801	45	YSC9186	YSC9186 Saccharomyc
24	92.8	22.6	16854	1	AE001722	AE001722 Thermotog
25	91.2	22.2	14168	1	AE001281	AE001281 Chlamydia
26	89.8	22.9	10195	65	U32737	U32737 Haemophilus
27	88	21.5	9940	1	AE002201	AE002201 Chlamydia
28	88	21.5	10447	1	AE001620	AE001620 Chlamydia
29	86.6	21.1	39228	2	MLCB1788	MLCB1788 Chlamydia
30	85.8	20.9	36021	2	MTCY359	MTCY359 Mycobacte
31	82.4	20.1	2053	65	SYOPGD	SYOPGD Mycobacte
32	82.4	20.1	2263	2	SSGNDG	SSGNDG Synchococc
33	82.2	20.0	1536	82	HSU30255	HSU30255 Human phosp
34	82.2	20.0	6278	1	BACNTBL	BACNTBL Bacillus li
35	82.2	20.0	1610	7	AB006102	AB006102 Candida a
36	81.6	19.9	43325	45	SPBC660	SPBC660 S. pombe c
37	80.8	19.6	1335	2	EVU14465	EVU14465 Escherichia
38	80.4	19.6	13799	1	AE002359	AE002359 Treponema
39	80.2	19.1	14838	1	AE001213	AE001213 Actinobacil
40	78.4	18.9	10903	2	D88189	D88189 Klebsiella
41	77.4	18.8	1335	2	KPU14463	KPU14463 Synchocyst
42	77.2	18.7	125469	65	SYCSLRA	SYCSLRA Synchocyst
43	76.8	18.3	10548	1	AE004417	AE004417 Vibrio ch
44	75.2	18.0	1335	2	CAU14426	CAU14426 Citrobacter
45	74	18.0	1335	2	CAU14426	CAU14426 Citrobacter

ALIGNMENTS

RESULT	1	AF061838	1686 bp	mRNA	PLN	08-JUN-2000
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						


```

/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA93016.1"
/db_xref="GI:7363272"
/translation="MPEADQDSDPADPTSPRPPLPGADYSKRGCTDMRAPLVSGSRE
RGDAGTAPLVAQCGCGGAGGAGTGHGSKREGAAGLGAASPRGARRRRRAAT
DGGDRDGDGRRRRHIAATMRGAARARRGEGVLTGVRATPEETTINGDAAT
ELQDTLREEMREVRGREGVHGERPWPPTAAAMEGLTGTRTTAFDGGIWRNRGRRAA
SGRRSGKGGGNDRGHGRKWKKEGKEGCEAL"
join(19096..19172,20123..20329,20389..20395)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA93017.1"
/db_xref="GI:7363273"
/translation="MOIEPTMHQCLVOWQCDKIEIIPANRPOSSRRRLGTAAPTRP
DILPSPPESSRRPTLPAASAAAPSRQPPRASPQSRHSISGSGALSRLSLRI"
21587..21769
/notes="5' LTR"
join(21843..22489,22572..23072,23168..24293)
/notes="Similar to Arabidopsis thaliana DNA chromosome 4,
BAC clone F17A8; putative protein. (AL049482)"
/codon_start=1
/protein_id="BAA93018.1"
/db_xref="GI:7363274"
/translation="MEKSSIALLFQKHEAKKNACDLPIPHVDEGSSIPAVNDAPDIED
IVYDEGCEETEEDTVDDAPPVVDVDEVSIEETKEESLPYDVIDDLEHDPGLRVFIS
SFANDQDAARRGYILKPCHLWAFNFPKRIYKDKDRFVTFHFKPIEYISVDKDA
TFPCVCLFGKESGKFTGQGWNNWAKALDKHVGTSDDHNFQAQERYNLVYKKGCL
RYLLROGLAFRGHDETESNNGNFLELLKWLGNENANAKVVLNNAPGNCILTSPIRI
QKQVRCSSWTSGLKELGDEHYATLADESSDISHKOLVCLRYVOKLGVGCVREF
LAVVHAGTSGLQKTAIQALISLTLTQIHQGYDGAASNMKGVEKCKWLPNLFQ
QVSLNLVIGVSTRHMDRVAQKLKALNGETAASGLNQMELALARPQDTRWGS
HYKTLILIMYDTIREVLTIGKQPTREDWPIIHAVLAFESFEFVNAHMLVIL
GYNEFNSLQKQDDIVNAMSVLGAKKKQOMRNGWEGFLGYSFCISYIDIP
AMDAYVPHGRSHRFYQYTDHRYREVYIGVIERHELENRFDVSEMLLCLMSA
FNPTDPSADQAKILASFPKDIKESLMLKLELDQTYINDMEDIRHFKLNGIK
QLSKLVETKHDLYLALLLKLVLIPVATASVERVESAMNLVTKVRSNMSDKLL
NCLVTFIERDMYMKKI"
26421..26603
/notes="3' LTR"
complement(join(29217..30334,31321..31609,31701..31862))
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA93019.1"
/db_xref="GI:7363275"
/translation="MYQIKIKKKRDEVLVATRLGERKVESLQALRMDGRVSGDDADGE
RRPLRRRCRQCGPKSVGLDMLGLPWITSLTHPPPLASTPTGRAAAAAVATAGL
VSSPAHPCKLRVSRWSTGAAGSGSGEDHGRRRRPPRRSSPPLSAKGLMDMLHQI
TALLSSODLILAFSGTCSRWAALSSPSTVTFPPHLKLPIDIPNSHPHCSFRVTL
LYKRWQLGDSKRTLSLRCSAPONTNRMRYLGCYGLFYSYENCLVDMVTGAK
VQPKLQSGAKNEYTYGLITPLNLPISHLILLCSRSSIFVQVGTNWSRHPFGGERI
LOVLFGEEFAMDFHRLHMRFPAPQLSMQEVGVWGMEEFVGVHFKPMLVISGDM
LMLDLVSIHSGYGPFTQVFLDFSAQAKWMEKLENSALFVSLDRNRNPTFCT
NPERWGGKSNICYVAKPTSEDSDEPWTAVELQPIPGATHCVYPYSHPLLRTEGHSOLE
YLWLPLSPFINGVQ"
32671..33849
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAA93020.1"
/db_xref="GI:7363276"
/translation="MQSPSKVSEASGAPPATSVIEGMAELPEGLLHLSIALLGSFLD
LIAFTGCHSRRAPFSSYPKSNFTILPPLLVNPNVRKAPVSSNGHRLSRSEVID
LANRFTPLRCOIPOTLORMHFAGSGHGLICCRGVCYLVVDVFTGAEVSPRPFPSE
NCDEYYCGILTAPTISNSHLIISTOSLFDMPVSDSWSELKLPVNRVDOIVVEFNG
QLIAVLESYTLQAPILRLKLEKTLIWNDMNECPYMRPWFVYVCGDMLIVDHIISF
SFGAPVLYRPLDMSTPAKWEVKLENNALFEGDARSPFSPKRNPRWGRSNC
LYYAHYSQPLSLHGLGDDADAVMDPNTDNLVFKRNVYRQLQALWVYPSMFYSAGDGO"
complement(join(34717..34935,35401..35912,36242..36363,
36470..36542,36626..36754,36936..37247,37336..38617,
38724..39212))
/notes="Similar to Arabidopsis thaliana DNA chromosome V,
BAC clone TM021B04; N. tabacum membrane-associated

```

```

salt-inducible protein. (AF007271)"
/codon_start=1
/protein_id="BAA93021.1"
/db_xref="GI:7363277"
/translation="MAAATAGAAASTAVSITCSSYSYEDDCAATSMLSGTSSTSSSS
SOROPYRLLHDEAQRLLRRRGSGAHTPRVRRTPDQMAHYVDEDDRAGHVRGRH
VVAAYRAARATASCSSPSADMRFAKVFALTFREMFCEHLHGRQARDFFAWM
KLQCLPEPSVAVYRILLRLYQGVKVAEVLFLMOACCEPDVACAGTLLCAARW
KLMNDLVAAVRRDIPVSIISVFNWSSLOKQKLGKVIHLWEQMLPANYNQF
TIVYIGSYAKEMLEAMDAFGMKRRRFRPEEATYSLLISLCAKHGKEGREALGYD
EMKVISVPSNYTCASVITLYIKNEDYSKALSLFSEMEQNKIYVDEVIYGLIYIKG
LGLYEDQRMFEEDKAGLLSDEQTYVMAQVHNNVQNYDRALQVLDAMRARNKPSQ
FYSALLRCHVAKEDVDAEDTFRALSNYGPDPVFCNDRLLRLYMRGLHLDKARALIL
KMRKLEQDFEDLCVTVLEVCCITSINKDNTLTVIQNESSSKVLNPTDSTLSNM
LKSLLDPGGLSVQLIMKFAFGSDDEAKFLYHELTGAKPDDTAIATLIVQYQ
AQOUEQAKLUFETASTFPVGGSVYNWADALCRCKGTEEARLFLMELIDQHGNGDAV
TISLIVTLTKQERFQEAENIYFCLHDEAELDTVYNTFIKSMESGLDNVFLIYVS
IYDRMISGIPRSMQTFNIMISYVGGGKLEKAVEMFSAQELGLPIDEKTYNMLSF
YKAGLHNEAETIFQEMKNNHVPDSHTYLAIRAYTEGKYSKAEARAIQMLRNSMT
PSCYHNLHLSAFLKEGIDEAORMYNOBEAGIPADLACRTMRMHLHOGYVDDGI
LFFETACRLLPDPSFILSAAPLHYHSGRSEFAGDVLDAINMSGASFLRNKLVGSKLE
QIQPKWMSIVKSGGLFPKAGHGVLPKQKSPWCSSVQVVARSEVDAAQRFGRDA
AVDHVQRYDSSTCD"
complement(join(40023..40413,40774..40971,41292..41464,
41783..42027,42111..42288,42719..42901,43197..43301,
44137..44274,44361..45353))
/notes="ESTs C74776(F51022),C261123(C116681) correspond to a
region of the predicted gene.
Similar to Arabidopsis thaliana cultivar Landsberg
extra-large G-protein (AF060942)"
/codon_start=1
/protein_id="BAA93022.1"
/db_xref="GI:7363278"
/translation="MAGAGGSSNWEEMRMLPPGTTIPEANLDYSTALEVDGP
PVSYELPRIDVDLPATPAOPVSPPLVPGNSGVAVPVVRPMPVHRQODHARAE
PPVAAQRRRRSSSVSDAPQNEGFSDDDSCSVSQAHNFGQRGGRGTAQGRER
AQVVTFTYEDSRYESKEDFVSQYVATTKRKRKRTCSRCKARKKRSKESCLVCDR
RPSYCVLVRMGSMPEGRKCTICQIPIDESKSKLGKSRILSLPLSVRLQILKA
BKECQANQLRPEQLIINGPLNDEMASILSCQRPQPKLKPGRYWDKESGLWKEGE"
Query Match 79.5%; Score 326; DB 8; Length 150120;
Best Local Similarity 88.5%; Pred. No. 7,6e-58;
Matches 353; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 12 qcgtcatggggcagacaccttgccctcaacattgcagagaaagggtcccccattctgtg 71
Db 50482 GCGGTCTATGGGGCAGAACCTTGCCCTCAACATTGCAGAGAAAGGGTTCCTATCTGTC 50541
Qy 72 tacaacaggagcaacctccaggctggacgagaccgtgcagcgtgccaaagcagaagaac 131
Db 50542 TACAACAGGACGACTTCTAAGGTGTGATGAGACCGCTTCAGCGCCCAAGGTAGAGGAAC 50601
Qy 132 cttcccgctacggctccatgaccgcgctctcttgaaagtcattcagagaccagc 191
Db 50602 CTTCTCTGTGACGGGTTCATGACCCCTGCACTCTTGTGACTCCATTCAGAGCCACGT 50661
Qy 192 gtggtgatcatgctcgtcaaggcgcgcgcagtgtagccagaccattcgcagcgtcga 251
Db 50662 GTTGTGATCATGCTTGTCAAGGCTGTGTGCACAGTGACCAAGGTAGCACTTGTGCA 50721
Qy 252 gctcaattgagacagggcagctcatcatcgatggggggaacagtggtacgagaacacg 311
Db 50722 GCACACTTGGACGAGGTGATGATTATTATTGGAGGAATGAGTGGTACGAGAACACT 50781
Qy 312 gagaggaggaagggcagtgagagcgcgcctnctgtatcttgagcatgggtgtctct 371
Db 50782 GAGAGGAGGAGAGGCAATGGAGACGGTGGCTCTCTATCTTGGTGGTGGTGTTC 50841
Qy 372 ggaggaagaggggtgcccacaacggccctccttgatg 410
Db 50842 GGAGGAGGAGGAGGTGCCCGCAATGGCCCTCCTCTATCTTGGTGGTGGTGTTC 50880

```

```

RESULT 4
AF037037 2392 bp DNA PLN 26-NOV-1998
LOCUS Zea mays 6-phosphogluconate dehydrogenase isoenzyme A gene, partial
DEFINITION
ACCESSION AF037037
VERSION AF037037.1 GI:3925238
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
REFERENCE
AUTHORS Padegimas, L.S. and Reichert, N.A.
TITLE Isolation of genes and regulatory sequences implicated in
hypersensitive response from Zea mays nematode-resistant line MP307
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2392)
AUTHORS Padegimas, L.S. and Reichert, N.A.
TITLE Direct Submission
JOURNAL Submitted (06-DEC-1997) Department of Plant and Soil Sciences,
Mississippi State University, Box 9555, Mississippi State, MS
39762, USA
FEATURES
Source Location/Qualifiers
1..2392
/organism="Zea mays"
/db_xref="taxon:4577"
1858..>2392
/note="6pgh(A)"
/codon_start=1
/product="6-phosphogluconate dehydrogenase isoenzyme A"
/protein_id="AAC79957.1"
/db_xref="GI:3925239"
/translation="MALTRIGLAGLVGMGNLALNIAEKPPISVYNTTTSKVDETVQ
RAKVEGNLPVFGFHPDPASFVSSIOKPRVIMLVKAGAPVDQTIATLAHLDQGDICVD
GGNEWENTREKAMEERGLLYLGMVSGEGEARGNPSLMPGGSFEAYKYIEDILL
KVAQVPDPSGFCVYIGK"
BASE COUNT 642 a 529 c 546 g 668 t 7 others
ORIGIN
Query Match 79.0%; Score 323.8; DB 7; Length 2392;
Best Local Similarity 88.0%; Pred. No. 5.4e-57;
Matches 352; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 11 cgcggtcatggtgagacacattgcctcaacattgcagagaaagggttccccatctgt 70
DB 1890 CGCGGTATGGGACAGAACCTTGCCCTCAACATCGCGGAGAAAGGTTCCCATCTCGT 1949
QY 71 gtacaacagagaaacctccaagggtgacgagaccgtgcagcgtgcgaaggcagaagaaa 130
DB 1950 CTACAACAGAGAACCTTCAAGGTTGATGAGACCGCTGACGCGTGCAGAGGAA 2009
QY 131 ccttcccgtaagcgttccatgaccccgctccttctggaagtcattcagaagccagc 190
DB 2010 CCTCCAGTGTGTTGGTTTCCACAGCCCCCGCTCTTCGTGAGCTCCATCCAGAGCCCG 2069
QY 191 ggtggtatctgctcgtcaaggcgcgcagtgccagtgaccagacattcgcgcgtcgc 250
DB 2070 TGTGCGTATGATCTGCTCAAGGCTGGGCGCGCGGTGACGACGACATTCGACGCTCGC 2129
QY 251 agctcacttggagcagcgactgcattcagatggggggaacagagtggtcagaacac 310
DB 2130 GCGCACCTTGATCAGGGGAGCTGATGCTGATGTTGGCAACGAGTGATGACGAGAAC 2189
QY 311 ggaagagagagagagggccattgagagcgcgccctnctgtatcttgcatgggtgctc 370
DB 2190 GGAGAGAGAGGAGAGGCGGATGAGAGAGCGGGGCTCTATATCTTGCAATGGCGTCTC 2249
QY 371 tgagagaaaggaggtgcccgaacgagccgctccttgatg 410
DB 2250 CGAGAGAGAGGAGGGTGGCCCGCAATGGCCGCTCTTGATG 2289

```

RESULT 5

```

AF061837 1818 bp mRNA PLN 08-JUN-2000
LOCUS Zea mays strain W64A x A182E putative cytosolic 6-phosphogluconate
DEFINITION dehydrogenase (pdh1) mRNA, complete cds.
ACCESSION AF061837
VERSION AF061837.1 GI:3342799
KEYWORDS
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
REFERENCE
AUTHORS Redinbaugh, M.G. and Campbell, W.H.
TITLE Nitrate regulation of the oxidative pentose phosphate pathway in
maize (Zea mays L.) root plastids: induction of 6-phosphogluconate
dehydrogenase activity, protein and transcript levels
JOURNAL Plant Sci. 134 (2), 129-140 (1998)
REFERENCE 2 (bases 1 to 1818)
AUTHORS Redinbaugh, M.G. and Bailey-Serres, J.
TITLE Characterization of cDNAs encoding cytosolic and a putative
plastidic 6-phosphogluconate dehydrogenase
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1818)
AUTHORS Redinbaugh, M.G., Campbell, W.H. and Bailey-Serres, J.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1998) USDA, ARS/Dept. Plant Path., OARDC/OSU,
1680 Madison Ave., Wooster, OH 44691, USA
FEATURES
Source Location/Qualifiers
1..1818
/organism="Zea mays"
/strain="W64A x A182E"
/db_xref="taxon:4577"
1..1818
/genes="pdh1"
/genes="pdh1"
69..1523
/note="NADPH producing dehydrogenase of the oxidative
pentose phosphate pathway."
/codon_start=1
/product="putative cytosolic 6-phosphogluconate
dehydrogenase"
/protein_id="AAC27702.1"
/db_xref="GI:3342800"
/translation="MALTRIGLAGLVGMGNLALNIAEKPPISVYNTTTSKVDETVQ
RAKVEGNLPVFGFHPDPASFVSSIOKPRVIMLVKAGAPVDQTIATLAHLDQGDICVD
GGNEWENTREKAMEERGLLYLGMVSGEGEARGNPSLMPGGSFEAYKYIEDILL
KVAQVPDPSGFCVYIGKGGSGNFVKMVHNGIEYDGMQIAEAYDVLKSGYKLTNSEL
HOVSEWNGKGLLSFLIEITADIFGDKDGEGLYVDKLDKMGKTGKWTQQAAL
LSVAAPTLEASLDSRFLSGLDKDERVEASKIPQGDYTGPLPDKAQLTEDYRQALYASK
ICSYAAGNNIIKAKSSSEKGLNLGELARIWKGCCIIRALFLDKKAYDRNPLANL
LVDPFAQEIIDROAARVRVCLINNGVSTPGMSASLAYFDSYRRDLPLANLVQQR
DYFGAHTYVERVDMPVLTSPRLRETTTSALTREKV"
BASE COUNT 418 a 474 c 520 g 406 t
ORIGIN
Query Match 78.2%; Score 320.6; DB 7; Length 1818;
Best Local Similarity 87.5%; Pred. No. 2.6e-56;
Matches 350; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 11 cgcggtcatggtgagacacattgcctcaacattgcagagaaagggttccccatctgt 70
DB 101 CGCGGTATGGGACAGAACCTTGCCCTCAACATCGCGGAGAAAGGTTCCCATCTCGT 160
QY 71 gtacaacagagaaacctccaagggtgacgagaccgtgcagcgtgcgaaggcagaagaaa 130
DB 161 CTACAACAGAGAGAACCTTCCAGGTTGATGAGACCGCTGACGCGTGCAGAGGTCGAAGAAA 220
QY 131 ccttcccgtaagcgttccatgaccccgctccttctggaagtcattcagaagccagc 190
DB 221 CCTCCCGTGTGTTGGTTTCCACAGCCCCCGCTCTCTGAGCTCCATCCAGAGCCCG 280

```

Qy	191	ggtggtgatcatgctcgtcaagcgccgagcgccagtgtagaccagaccatgcgacgcctgc	250
Db	281	TGTCGTATCATGCTCGTCAAGGCTGGGGCGCGGTGGACGACACCATTCGCCACGCTCG	340
Qy	251	agctcaacttgagcagcgccagctgcatcatcgatgggggaaacagagtggtacagaacac	310
Db	341	GGCGACACTTATCAGGGGAGCTATCGTCGATGGTGGCAACGAGTGGTATGAGAACAC	400
Qy	311	ggagagagagagagccatggagagcgccgctnctgtatcttggcatgggtgtctc	370
Db	401	GGAGAGAGAGAGAGAGGCGATGGAGGCGGGCTCTTATCTTGGCATGGGCTCTC	460
Qy	371	tggaggaagaggggtgcccgcaacgcccgcctcttgatg	410
Db	461	CGGAGGAGAGAGGCTGCCCGCAATGGCCGCTCTTGATG	500
RESULT	6		
LOCUS	AF037029	645 bp mRNA	26-NOV-1998
DEFINITION	Zea mays 6-phosphogluconate dehydrogenase isoenzyme A mRNA, partial cds.		
ACCESSION	AF037029		
VERSION	AF037029.1	GI:3925222	
KEYWORDS	Zea mays.		
SOURCE	Zea mays.		
ORGANISM	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.		
REFERENCE	1 (bases 1 to 645)		
AUTHORS	Padegimas,L.S. and Reichert,N.A.		
TITLE	Isolation of genes and regulatory sequences implicated in hypersensitive response from Zea mays line MP307		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 645)		
AUTHORS	Padegimas,L.S. and Reichert,N.A.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-DEC-1997) Department of Plant and Soil Sciences, Mississippi State University, Box 9555, Mississippi State, MS 39762, USA		
FEATURES	Location/Qualifiers		
Source	1..645		
CDS	/organism="Zea mays" /db_xref="taxon:4577" 111..>645 /note="6pgdh(A)" /codon_start=1 /product="6-phosphogluconate dehydrogenase isoenzyme A" /protein_id="AAC79949.1" /db_xref="GI:3925223" /translation="MALTRIGLAGLVNMGONLALNAEKFPISVDRRTTSKVDETQV RAKVEGNLPVGFHPDPAFVSIOKPRVIMLVKAGAPVDQTATLAHLDDQGCIVD GNEWYENTERREKAMEERGLLYLGMVSGGEBGARDGPSLMPGGSFEAYKIIEDILL KVAAQVPDPSGPCVTYIGK"		
BASE COUNT	131 a	193 c	204 g
ORIGIN	117 t		
Query Match	77.8%;	Score 319;	DB 7; Length 645;
Best Local Similarity	87.2%;	Pred. No. 7e-56;	
Matches	349; Conservative	0; Mismatches	51; Indels
Qy	11	cgcggctatgggcagacacttccctcaacttcagatgcagagaagggttccccatctctgt	70
Db	143	CGCGGTATGGGACAGAACCTTCCTCAACATCGCGGAGAAAGGTTCCCATCTCGT	202
Qy	71	gtacacagagacacacctccaagtgtgacgagaccgtgcagcgtgcagagaagga	130
Db	203	CTACGACAGACAACTTCCAAAGTTGATGAGACCGTGCAGCGGCCCAAGGTGGAAGAAA	262
Qy	131	ccttcctcgtactacggttcctcatgaccgcgctcttcttgagatccattcagaagccacg	190

Db	263	CCTCCCGCTGTTGGTTTCCACGACCCCGCTCTCTTCGTGAGCTCCATCCAGAACGCCCG	322
Qy	191	gggtggtatcatgctcgtcaagcgccgagcgccagtgtagaccagaccatgcgacgcctgc	250
Db	323	TGTCGTATCATGCTCGTCAAGGCTGGGGCGCGGTGGACGACACCATTCGCCACGCTCG	382
Qy	251	agctcaacttgagcagcgccagctgcatcatcgatgggggaaacagagtggtacagaacac	310
Db	383	GGCGACACTTATCAGGGGAGCTATCGTCGATGGTGGCAACGAGTGGTATGAGAACAC	442
Qy	311	ggagagagagagagccatggagagcgccgctnctgtatcttggcatgggtgtctc	370
Db	443	GGAGAGAGAGAGAGGCGATGGAGGCGGGCTCTTATCTTGGCATGGGCTCTC	502
Qy	371	tggaggaagaggggtgcccgcaacgcccgcctcttgatg	410
Db	503	CGGAGGAGAGAGGCTGCCCGCAATGGCCGCTCTTGATG	542
RESULT	7		
LOCUS	AC068900	96240 bp DNA	HTG
DEFINITION	Arabidopsis thaliana chromosome III clone IGF-F11A12, *** SEQUENCING IN PROGRESS ***, 17 unordered pieces.		
ACCESSION	AC068900		
VERSION	AC068900.2	GI:8192588	
KEYWORDS	HTG; HTGS, PHASE1.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	1 (bases 1 to 96240)		
AUTHORS	Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.		
TITLE	Arabidopsis thaliana 'IGF' BAC 'F11A12' genomic sequence near marker 'mi74'		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 96240)		
AUTHORS	Lin,X. and Kaul,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-MAY-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@igr.org		
COMMENT	On Jun 2, 2000 this sequence version replaced gi:7769972. * NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.		
	1	22087: contig of 22087 bp in length	
	22088	22137: gap of unknown length	
	22138	25099: contig of 2962 bp in length	
	25100	25149: gap of unknown length	
	25150	27221: contig of 2072 bp in length	
	27222	27271: gap of unknown length	
	27272	29774: contig of 2503 bp in length	
	29775	29824: gap of unknown length	
	29825	34383: contig of 4559 bp in length	
	34384	34433: gap of unknown length	
	37434	37484: gap of unknown length	
	37435	41731: contig of 4247 bp in length	
	37485	41781: gap of unknown length	
	41732	43856: contig of 2075 bp in length	
	41782	43857: gap of unknown length	
	43907	46860: contig of 2954 bp in length	
	46861	46910: gap of unknown length	
	46911	50824: contig of 3914 bp in length	
	50825	50874: gap of unknown length	

* 50875 56713: contig of 5839 bp in length
 * 56714 56763: gap of unknown length
 * 61235 61235: contig of 4472 bp in length
 * 61236 61285: gap of unknown length
 * 61286 66859: contig of 5574 bp in length
 * 66910 73521: contig of 6612 bp in length
 * 73522 73571: gap of unknown length
 * 73572 78793: contig of 5222 bp in length
 * 78794 78843: gap of unknown length
 * 78844 83051: contig of 4208 bp in length
 * 83052 83101: gap of unknown length
 * 83102 96240: contig of 13139 bp in length.

FEATURES

source

Location/Qualifiers
 1..96240
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="III"
 /clone="IGF-F11A12"
 BASE COUNT 28496 a 18610 c 20275 g 28013 t 846 others
 ORIGIN

Query Match 63.9%; Score 262; DB 56; Length 96240;
 Best Local Similarity 78.4%; Pred. No. 1e-44;
 Matches 313; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 12 gcgggtcatggggcagaaccttgcctcaacattgacagaaggggtcccatctctgtg 71
 Db 13068 GCTGTGATGGGTCAGAACTAGCTCTCAACATTGCAGAAAGGCTTCCCAATCTCTGT 13127
 QY 72 tacaacaggacaacctccaaggtggacgagaccgtgcagctgcgaagggcagaaggaac 131
 Db 13128 TACACAGAACTACTTCCAAAGTTGATGAGACCGTCGACGAGCCAAAGAGGCAAT 13187
 QY 132 ctcccgctacgcttcattgacacccgcgctccttggtaagtcattccatcagaagccag 191
 Db 13198 CTTCCTCTTTACGGCTTCATGACCCCGAGTCTTTTCTCAAGTCCATTCAGAGCCACGC 13247
 QY 192 gtggtgatcatgctgcgaagccgagccagctgacagaccgtggggaacagtggtacgagaacac 311
 Db 13248 GTTATTATCATGCTGTGTAAGCTGGTCTCCCGTTCACGAGACCATCAAGACCTCTCT 13307
 QY 252 gctcacttgagcaggcgacgtcatcgtatgagtggtgggaacagtggtacgagaacac 311
 Db 13308 GCTTATTGGAGAAAGGTGATGATGCTGATGATGATGATGATGATGATGATGATGATGAT 13367
 QY 312 gagagggagagagccatggagagcgcgacctctgtatcttgcatgggtgtctct 371
 Db 13368 GAGAGGAGAGAAAGCGGTGGCGGAGAACGGTTCTCTACCTTGGAAATGGGAGTTTCT 13427
 QY 372 ggaggaaggagggtgcccgaacgcccgcctctcttgatg 410
 Db 13428 GGTGGTGAAGAAGGTGCTGCTAATGGCCCATCATGATG 13466

RESULT 8

AB007907
 LOCUS
 DEFINITION
 Glycine max mRNA for 6-phosphogluconate dehydrogenase, complete cds.
 AB007907
 ACCESSION
 AB007907.1 GI:2529228
 VERSION
 6-phosphogluconate dehydrogenase; gnd.
 KEYWORDS
 Glycine max root nodules cDNA to mra.
 SOURCE
 Glycine max
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Fabales; Fabaceae; Papilionoideae; Glycine.
 REFERENCE
 1 (bases 1 to 1667)
 Katsurada, A.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (07-OCT-1997) to the DDBJ/EMBL/GenBank databases. Akihiko

Katsurada, Tezukayama Gakuin College, Dept. of Nutrition; 4-2-2
 Harumidai, Sakai, Osaka 590-01, Japan
 (E-mail: LDD00547@niftyserve.or.jp, Tel: +81-722-96-1331,
 Fax: +81-722-92-2135)
 2 (bases 1 to 1667)
 Katsurada, A.
 TITLE
 Molecular cloning of 6-phosphogluconate dehydrogenase full length
 cDNA from soybean
 JOURNAL
 Published Only in DataBase (1997) In press
 FEATURES
 Location/Qualifiers
 1..1667
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /tissue_type="root nodules"
 55..1590
 /gene="gnd"
 55..1590
 /gene="gnd"
 /EC_number="1.1.1.44"
 /codon_start=1
 /product="6-phosphogluconate dehydrogenase"
 /protein_id="BAA22812.1"
 /db_xref="GI:2529229"
 /translation="MAQSTGIGLAGLAVMGONLALNAEKGFPISVYNRTTSKVDET
 VRKQEGNLPVIGYHDEPAFVHSIQPRVIMLVKAGAPVDQITKISAYMEKGDGI
 IDGNEWENTERREREKVAELGLYLGLVSGGEGARGNSLMPGGFSAFKYLEDI
 LLKVAQVPDSCPVYIGKGGSNFVKMIHNGIEYGMOLIAEAYDLVSKVGLSNE
 ELQSVFSEWNKCELLSFLIEITADIFGDKDKDGLVLDKLDTKMGKGTQWTVQA
 AELSIAAPTIEASLDARLSGLKEERVEAAKVFSGGIGDIVDHOVDKLDIDVRK
 ALYAKICSYAGGMNLIIRAKSTEKWDKLGELRIWKGGCIIRAFIDRLTKQAYERN
 PNLALLVDPEFAKEIIDYQSARRVVCILAINSGISTFGMSASLAYFDYTRRLPAN
 LVQARDYFGAHTYERVDIEGSHTEFKLAKQSKLDYCIISANQDFPNKCNIFCSDC
 MLS"
 1667

polyA_site

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 293;

Conservative

Mismatches 106;

Indels 0;

Gaps 0;

QY 12

Db 94

QY 72

Db 154

QY 132

Db 214

QY 192

Db 274

QY 252

Db 334

QY 312

Db 394

QY 372

Db 454

RESULT 9

[illegible]

RSKFDKSFSSSLNCGDVYTPYMDKLEERTFTVTYSIVTPLDNNAFQVATALL
KGCIVQLSDCCTGDRORYPCPLHALAVCKKLFPNLOVYDVCYTLERLKVAT
IFSHVEMSAWPAWGPRLPPVIPPSPPTTYSVREKVMWCHLNGIKGAD
VYEGESILKKTIMKRTALSDRLYLQFLDKQSEFKIFGYIPVDVSDPLETYV
HPTVLNDSNNGSAGILVKNVNSGALKPCLLPLRIWDDHMHGLGCFKDRDELKK
AUMWCIIRBNCIVRETEKEMVTECVKWKCKWSLRAARMEHGLVETTKYTGPHTC
SHYENDPSEFADEIERVRIQPTLSIAELKWKWKEKTYGELQTSKMRDGLKLEIK
RVFEDQSFRVMPKLISAFSHSNGLLVDQXLPNDFPASFVGFWSFSQISGFQ
HCRPLIVDTKSLNGYOLKLMIAAGVADKEFFPLAFATKEVTSQWNRWFFKIRE
KVRQKDLCLISSPLRDIVAVNPGSLWQEPWAHFKCLNHLRSQFLGVRFDYNLES
LVQAGCTNOKBEFDSYMDIKEKNPEAKWLDOPRHKWALAHSGLYRYEIDRE
ALFACVCGPYCTVAMTGVMLEDELRSDFKSLSIYSSINRGVYVTEPMEIDLEE
FMTDSIPVITQLERSFVSSSEKEEINVLNVSTCTCRKFSQYKFKFCLHALAVEE
KLKINPLQVDECTVEQYCKTAATFSPVDPVAAWPCDRCVPTLFPFSSQJSPNT
complement(join(15674..15817,15904..16008,16075..16196,
17033..17102))
/note="similar to ubiquitin-conjugating enzyme E2-17 kd
spiP35135; similar TO ESTs gb|184048, gb|745251,
gb|A100555, gb|AA041165, gb|T21697, emb|429145,
gb|T20867"
/codon_start=1
/evidence=not_experimental
/product="F22C12.2"
/protein_id="AAF24583.1"
/db_xref="GI:6692118"

CDS

/translation="MASKRLIKELKLDQDPPTSCSAVAEDFMHQAOTMGPSDSPYS
GGVFLVTHFPDPYPPKPPKAVATKVFHPNVSNGSICLDILKEQWSPALTISKVLL
SIGSLTDPNPDPLVPEAHMYKTDRAKYESTARSWTOKYAM"
complement(join(18031..18617,18685..19861))
/note="putative receptor kinase emb|CAA23040.1"
/codon_start=1
/evidence=not_experimental
/product="F22C12.3"
/protein_id="AAF24582.1"
/db_xref="GI:6692117"

CDS

/translation="MOITLFFSLIICFVLISQTLDDKKALLHFLSSFSRRLHN
QSDPVCNWSGVTCNDGRIIVSVRPVAVGNLIPPTISRLSSKFLSLRKNHFTG
DFSPDFTSLKPHLYLOHNLSPILAIIFSELNKLKVLDDLSNNGPNSIPTSLSLGT
SLQVNLNANSFGEIPNLHLPKLSQINLNKILGTPKSLQRFSSAFSGNNLTER
KKQRKTPFGLSLAFLLLSAACVLCVSLGSFIMTCFKTRISGRLKLRDSSSPGN
WTSRDDTEGGKIFEGGRNHLFDLDDLSSEAIVLGAGFTTYKVMEDMSTVVV
KRKEVYHVPDFEQMEIICMIRHENVAELKAYYSKDDKLVYIYNGHSFELIH
GNRGYHRVLDLARIATGAARGLAKTHEGFTGNKSNIFDLDCYQICIGDV
GLTIIKMSLPQTLCTSCYHAPETIDTRSTQFSDVYSGVVLLELTGKSPVSOAEL
VPTGGMNDLADSIWRSVVAKEWTGEVDFMELLSQSGGFEEMVEMLIQGLACVALQKQ
ERPHIAQVLKIEDIRSDAE"
join(20892..20993,21210..21310,21520..21694,21780..21976,
22063..22162,22282..22387,23017..23183)
/note="similar to vacuolar ATPase gl|3600058"
/codon_start=1
/evidence=not_experimental
/product="F22C12.4"
/protein_id="AAF24559.1"
/db_xref="GI:6692094"

CDS

/translation="MNDADASIQQQWVFIROBAEAKANEISISSEEFNIEKQLIV
FAEKKIROEYKKEKQVDRKKIDYSMOLNASRIKVLQADDDIVNMAEAAKQLLK
VSQGFNHHHKKYKLLKDLIVQCLRLKEPVLRLCREDDIDIVESMLDDASEVC
KKAHAPETIIVDKDIFLPAPSDDDPHALSCAGGVVLASRDKIVGENTILARLEVA
FRNKLPEFCSGFLEKQVDPKVALRGQWLSMSDSNFIITKEKLRDAKSNMPTGRKCP
DPNGVEKSMCYSSCKTQFGMGSCQGHKNYMCCEY"
23882..25345
/note="putative 6-phosphogluconate dehydrogenase
db|JBA22812; similar to ESTs emb|F15308, emb|15307,
gb|T21894, gb|A199770.1"
/codon_start=1
/evidence=not_experimental
/product="F22C12.5"
/protein_id="AAF24560.1"
/db_xref="GI:6692095"

CDS

/translation="MESALSRIGLAGLVGMGNALNIAFKGPPIVYNRVTSKYDE
TLDRABEGNLPVSGVSPDFVLSIORPILIVKAGAPVDOTIDAFSEVPGDC
IIDGNGENYQNTERRISEAQKGLLLYLGWVSGGEGEARGNPSLMPGSGFAYDNKID

ILEKVAQVEDGPCVTYIGBGGSGNFVKMWHNGTEYGMOLISEAYDVKNVGLSNE
ELAEFTWNSGLESFLVETSDIFRVKDFGDELVDKILDKTGMKGTKWTVQQA
AELSVAAPTAASLDCRYLSGLDERENAAKVLREAGLKEIGSASSGDKRLVDDV
ROALYASKISYAGCMNLLRAKSLKSWNLNFGELARIWKGCCIIIRAVFLDRIKKAYQ
ANLVQASLVVDPEFAKEMVQQAARVRVGLAVSAGISITPMGCASLAFDITYRARLP
RNLVQASLDRLFAGHYERTDPGAYHTWTKLARKNH"
complement(join(25806..26028,26204..26373,26722..28161,
28379..28489,28868..29254))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F22C12.6"

CDS

Query Match 44.6%; Score 183; DB 7; Length 111222;
Best Local Similarity 66.0%; Pred. No. 1.5e-28;
Matches 264; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 11 CCGCGTGTGGGCGAGAACCTTGCCTCAACATTCAGAGAAAGGTTCCCATCTCAGT 70
DB 23923 CGCGGTGTGGGCGAGAACCTTGCCTTAAACATCGCGAGAAAGGTTCCCATCTCAGT 23982
QY 71 gtacacagagcaacctccacaggtgagacacgagcagcgtgccaagagcaagagaaa 130
DB 23983 CTACAATCGAACCACTTCCAAAGTCGAGAAACCTAGATCGAGCGCGCTCGAAGGAAA 24042
QY 131 cctcccgctacggttcctcagcgcgcgtcctttgtgaagtcattcagaagccacg 190
DB 24043 CTTCCCGAGTCTCGGCCAATATCTCTCCCGGAGATTCCTCTCGATCCACAGCGCTAG 24102
QY 191 ggtggtgatcgtcgtcgaagcggcgccagtcagcagcagcagcagcagcagcagc 250
DB 24103 ATCTCTCATCTCTCGTCAAAAGCTGGAGTCCCTGTGACCAACAATCGAGCGCTCTC 24162
QY 251 agtcaactggagcggcgagcagcagcagcagcagcagcagcagcagcagcagcagc 310
DB 24163 CGAATACATGGAGCGAGGAGATTCATCATCGAGTGGAAACGATGGTATCAGAACAC 24222
QY 311 ggagagagggagagggcagcagcagcagcagcagcagcagcagcagcagcagcagc 370
DB 24223 AGAGCGACGAGTCTCGAAAGCTGACGAAAGAGATTCCTCTATTAGTATGGAGTCTC 24282
QY 371 tgagagagggaggggtgcccgaagcggcgccctctgtatcttggcaggggtgtctc 410
DB 24283 CGCGCGGAGAGAGGAGTCTCGTAAACGGTCTCTTCGTTAATG 24322

RESULT 12

AB005233/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

AB005233 79837 bp DNA PLN 20-NOV-1999
Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MBK23,
complete sequence.
AB005233 GI:2264305
HTG.
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:MBK23.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (sites)
Sato, S., Kotani, H., Nakamura, Y., Kaneko, T., Asamizu, E., Fukami, M.,
Miyajima, N. and Tabata, S.
Structural analysis of Arabidopsis thaliana chromosome 5. I.
Sequence features of the 1.6 Mb regions covered by twenty
physically assigned P1 clones
DNA Res. 4 (3), 215-230 (1997)
97471969
2 (bases 1 to 79837)
Nakamura, Y.
Direct Submission
Submitted (02-JUL-1997) to the DDBJ/EMBL/GenBank databases.

Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail: ynakamu@kazusa.or.jp, Tel.: +81-438-52-3935, Fax: +81-438-52-3934)

[illegible]

FEATURES	source
tax_id:36323934	Location/Qualifiers
	l..79837
	/organism="Arabidopsis thal
	/strain="Columbia"
	/db_xref="taxon:3702"
	/chromosome="5"
	/clone="MBK23"
	/clone_lib="Mitsui pl"
BASE COUNT	25977 t 14362 c 13908 g 25590 t

Query Match	42.7%	Score 175;	DB 7;	Length 79837;
Best Local Similarity	64.8%;	Pred. No. 6.e-27;		
Matches 259;	Conservative 0;	Mismatches 141;	Indels 0;	Gaps 0;
QY	11	cqcggtcatggggcagaaccttggcctccaacattgagagaaaagggttccccatctgt	70	
Db	62024	CCGAGTCATGGGACAAACCTTCGCCTTTAAACATCGCGGATAAAGGATTCCTCAATTCGGT	61965	
QY	71	gtacaacaggacaacctccaaggtggagcagacgctgcagcgtgccaaaggcagaagaaa	130	
Db	61964	CTACAATCGAACCACTTCCAAAGTCGACGAACCTTAGATCGTGCCTCCACGACGAGAA	61905	
QY	131	ctctcccgctfacagggttccaatgaccgccggtctcttgtgaagtcattcagaagccacg	190	
Db	61904	ACTCCAGTCGCTGGTCAAACTACTCGCTCGCGATTTCCGTCTCGATCCAACGGCCTAG	61845	
QY	191	ggtggtgatcatgctcgtcaaggccgcgccgagttgaccagacaatcgagacgtcgc	250	
Db	61844	ATCCGTTTATCATCTTGTCAAAAGCGGTGTCCTCCGTTGACCAACCATCTCGTCTCTCTC	61785	
QY	251	agctcaattggagcaggcgactgcatactgatggggggaacagagtggtacgagaacac	310	
Db	61784	TGAATACATGGAGCGCTGGTGAATTGATATCGACGGTGGAAATAGTGGTATCAGAACAC	61725	
QY	311	ggagaggaggagagaagggccatggagagcgccctctctatcttgcattggatggtctc	370	
Db	61724	AGAGCGAGGAATCTGTCAAGCTGAGAGAAGAGATTGCTTTATTAGGTATGGGAGTCTC	61665	
QY	371	tggagaaaaggaggggtccccgcacggcccgctcttctgatg	410	
Db	61664	CGGTGGTGAAGAAGGAGCTCGTAATGGTCTCTCTCTTATG	61625	

RESULT 13

AC011913	AC011913	38587 bp	DNA	INV	14-MAY-2000
LOCUS	Leishmania major chromosome 35 clone L4407 strain Friedlin,				
DEFINITION	complete sequence.				
ACCESSION	AC011913				
VERSION	AC011913.2	GI:7658329			
KEYWORDS	HTG.				
SOURCE	Leishmania major.				
ORGANISM	Leishmania major.				

REFERENCE

leishmaniasis.
1 (bases 1 to 38587)
Myler, P.J., Sisk, E., Cawthra, J., Vogt, C., Robertson, L.,
McDonagh, P. and Stuart, K.
TITLE
JOURNAL
Direct Submission
Submitted (16-OCT-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
2 (bases 1 to 38587)
Myler, P.J., Sisk, E., Cawthra, J., Handley, F., Vogt, C., Robertson, L.,
McDonagh, P., Stuart, K. and Ivens, A.
TITLE
JOURNAL
Direct Submission
Submitted (28-APR-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA

REFERENCE
AUTHORS

ADHOKS	TITLE	COMMENT
	Submitted (14-MAY-2000) Seattle Biomedical Research Institute, 4 Direct Submission	
	McDonagh, P., Ivens, A., Nguyen, D., Munden, H. and Stuart, K.	
	Myler, P.J., Sisk, E., Cawthra, J., Handley, F., Vogt, C., Robertson, L., Wright, A., and Brown, J.	
	On April 28, 2000 this sequence version replaced gl:6056180.	
	NOTE: This sequence is still preliminary and may contain some errors, such as substitutions and framehifts. please regard all CDS sequences with some degree of caution. A finished version should be available within a few weeks.	

FEATURES

```

FEATURES
source
Location/Qualifiers
1. .38587
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="L4407"
620. .2380
/gene="L354.2"
/note="L4407.1; predicted using Glimmer, Testcode, and
CodonUsage; Codonusage marginal; Glimmer predicts start
codon at 3374"
620. .2380
/gene="L354.2"
/codon_start=1
/product="L354.2"
/protein_id="AAF70120.1"
/db_xref="GI:7801403"
/translation="MQLDWDPTRYRETLLYARELAVLLPSSWSELSEQSAKSSSSLS
HSSLPSLAFVKVKEAPLTLQMTAAAAVPTCHGSSVDSEMDYATVAVAARWTR
MLGMIVLRVPELLCTSGVLFLABEYEVGRYFVPLAWMLDRSIERLAARLCA
LADKRRVQVCGTGTCANDMAASFCGTLGKDEDEQACDEVCTAMEATORVAREQ
LRI5DWDLHPSSIPATVGAKLKLRAQOQSRHSESTGSLPRPVVHTNCLGLLHALT
SFRLPQPHQADHGAPLLTLTQNALLHIGYIDLLTSQSSVIERCTYLIPALLDYEC
LLFCAMLKESALQAQKEKELTQTQSLIKTLQSLQSLQSAQGMVRIEALLVDYKKS
KSSASSTPRLVYKVAHLLPAMPLVAQAAQHASDAVVPKPTHPGAADGCSASA
VMRSEAPFTVLVELEPTCDLLQRYPADNWLTRSLTGPDRQDQATSGSHIAREG
GILLSPSSRRQRRLVSQOQRQSAMDVRRREVVSFLDAEVSNCFAAAGSGAAASKKQKQ
AADAAYLATYLSAYGPHWL"

```

gene

```

gene
AACAAATATATATLTAAGTGGPHWL"
4545. .6464
/gene="L354.3"
/note="L4407.2: predicted using Glimmer, Testcode and
CodonUsage; ORF starts at 6990, Testcode and CodonUsage
predict 2nd ARG used as start codon"
4545. .6464
/gene="L354.3"
/codon_start=1

```

CDS

```

predict 2nd ATG used as start codon"
4545. .6464
/genes="L354.3"
/codon_start=1
/product="L354.3"
/protein_id="AAF70121.1"
/db_xref="GI:7801404"
/translation="MVLSPVITLRMVHVFSTPTSLHSRSSTSPDLALVDFVSSFP
ATTSTPISRSOLPPELMERVCDPVCGLIRLSLLSYAERFESKMGQHVHVRSRQURQ
ESAAPCRANAGAEVGYSEFVSPSPSGNAAAASKNSVDVQAVTELEQQLPFSPYLVP
VKALNITVPAEQQVQVRAVLDAIRLSPQHRWLTFHEHPAHCAHSTGADAEALTHVL
EKGVQLRNLRQDRAAAPMPQSPNTRMAGHEVQVPEWYRVARVLPTTATELLFSG
ELQEQTALLPPGRDVMHLLGAPALFELRYAAPTVDSGQARVPTLSAEGSSLAPV
CVKYDATPPSLCASVYVRELFREVPVPMAGAREESILRELETLTERRASAGLS
SQRRRRKLRQLAVLRNPTPYFDERVLAQHLFDLLPLRGDGHQSAIIGSLUPQPAVQ
AFONTELLQANDLFQSDVHVLGTLVRRAADPAQKRSIESVTEGTEILLHFSSYS
LRSDPQEGTIISSRLPRLRFLFTMQDIIAEVLLYPDKVEVLADRSPPSISLS
LSSSPHTSADERALRELATVRGRDRFLVPFRFVGEGWQAKLTKYTKQTKQDKAVSMR
AQRQSFSSHRRPARY"
7615. .8880
/genes="L354.4"
/notes="L4407.3; predicted using Glimmer, Testcode and
CodonUsage"
7615. .8880
/genes="L354.4"
/codon_start=1
/product="L354.4"
/protein_id="AAF70122.1"
/db_xref="GI:7801405"

```

gene
9804...10592
AVSGASGLCIHHYGGDRAALRAVAMLLKGD #
YRVDELDFVETQYLMRIQBHFNAASYGKT KFSLSATLSWLGDRNTFMANNMYFTIGAVS
EAAADGNPNPSPMRMSIDNENQWADDEGPHRIPTVTDGDMWVARVASLKPFRKL
LHSAQSFRTNEPLDAGCKHVKGLAMTSDFEYKKPHFPSPSPSPVAPASAKWAL
LRHVDGAKPYYVYGLYENQYQNRHFRSCSGSYNTGAFRVSASFGDQWTCSTPTGP
VCLPSLSLRANASVDVIRYDQRCQSSGTSYNTGAFRVSASFGDQWTCSTPTGP
/translation-"MDVAPAAVVPBESVWALFKQORLPANQOPIITLTPQHSALCLIAVAP

gene
9804...10592
/note="L354.5"
CodonUsage: scores for Testcode and CodonUsage are relatively in the middle of the Ops"

```

8804 . 10592
/gene="L354.5"
/codon_start=1
/product="L354.5"
/protein_id="AAF70123.1"
/db_xref="GI:7801406"
/translation="MALDIFDDADTFPELVRDARLDYMGRTMTACWELADQHTVF
RDIAKERTGEILGAVLAAVQICGALQVELSNPEVRAPSPACTALFSPQAQVAVG
DTRGRMLLEAGATQAVRSYVLSYCTPTPTCAQWRSLQFVFPQPLNAVSTAANG
DTPRFLDEAGATPAGDAQVATPSTLYRTLKFQSDVDDVAQLPDSEVDKGGDFAEKLE
LVVCCMTVGRVRAPMTPLPLDAILVRENKVKLQDFVGNSTLEW"
11412 . 13055

```

gene	11412..13055	/gene="L354.6"	/note="L4407.5; predicted using Glimmer, Testcode and CodonUsage"	11412..13055
gnc	11412..13055			

CDS

```

/codon_start=1
/product="L354.6"
/protein_id="AAF70124.1"
/db_xref="GI:7801407"
/translation="MGSGYIKRRQRQELLNDFYSALSSDFSTTSGQFSKLVTRIV
HYPSLPVLLVDDPAEYRGRNKDKISAHWGQKLLUSEIQLLSMTCRPEISY
YVAGAAPLTHLDFMFCRHTWLEIDPGQFDRAVLEPRANFLRNEFTNATAYG
INARLTVPALGVQYVNAVSVQDKKAIHQGLQVITLDVARTGIEDIPAMTEA
PLSPKGLLELAQVALERNKPLVTFIDKSGSVLPNFEDVAENMRAGCEWNILHG
EHSMLKPLRPYTSKSTKFGCGGAVRETLILREDGTVPYLRGDMLLPLWTPTSTEGR
VYPQAGQKDNVAHVEDQDFPFNSVRQGVPHNLVHAHDSLELDHHIDAAAEVHCLLT
YLRTIRPELTKATGLRVLKLVHANSITVAHLGITFEDAIRRDAIMHAKKSQNK
VSGHDDQAEQAEKSAASKYRAGNSAAADCAATSTAPVPPSEWQETTR
LVRI.SYNDNRKNSVYNSREESHKSTSGFWITKMPQA"

```

```

gene
14174. .17254
LAVYLNILDRKSYVGNKNNYKZHHZHLJGFWYLLNGKWR
/genes="L354.7"
/note="L407.6; predicted using Glimmer, Testcode and
CodonUsage, ORF starts at 16556, 2nd ATG predicted as
start_codon"

```

117254
 117255
 117256
 117257
 117258
 117259
 117260
 117261
 117262
 117263
 117264
 117265
 117266
 117267
 117268
 117269
 117270
 117271
 117272
 117273
 117274
 117275
 117276
 117277
 117278
 117279
 117280
 117281
 117282
 117283
 117284
 117285
 117286
 117287
 117288
 117289
 117290
 117291
 117292
 117293
 117294
 117295
 117296
 117297
 117298
 117299
 117300
 117301
 117302
 117303
 117304
 117305
 117306
 117307
 117308
 117309
 117310
 117311
 117312
 117313
 117314
 117315
 117316
 117317
 117318
 117319
 117320
 117321
 117322
 117323
 117324
 117325
 117326
 117327
 117328
 117329
 117330
 117331
 117332
 117333
 117334
 117335
 117336
 117337
 117338
 117339
 117340
 117341
 117342
 117343
 117344
 117345
 117346
 117347
 117348
 117349
 117350
 117351
 117352
 117353
 117354
 117355
 117356
 117357
 117358
 117359
 117360
 117361
 117362
 117363
 117364
 117365
 117366
 117367
 117368
 117369
 117370
 117371
 117372
 117373
 117374
 117375
 117376
 117377
 117378
 117379
 117380
 117381
 117382
 117383
 117384
 117385
 117386
 117387
 117388
 117389
 117390
 117391
 117392
 117393
 117394
 117395
 117396
 117397
 117398
 117399
 117400
 117401
 117402
 117403
 117404
 117405
 117406
 117407
 117408
 117409
 117410
 117411
 117412
 117413
 117414
 117415
 117416
 117417
 117418
 117419
 117420
 117421
 117422
 117423
 117424
 117425
 117426
 117427
 117428
 117429
 117430
 117431
 117432
 117433
 117434
 117435
 117436
 117437
 117438
 117439
 117440
 117441
 117442
 117443
 117444
 117445
 117446
 117447
 117448
 117449
 117450
 117451
 117452
 117453
 117454
 117455
 117456
 117457
 117458
 117459
 117460
 117461
 117462
 117463
 117464
 117465
 117466
 117467
 117468
 117469
 117470
 117471
 117472
 117473
 117474
 117475
 117476
 117477
 117478
 117479
 117480
 117481
 117482
 117483
 117484
 117485
 117486
 117487
 117488
 117489
 117490
 117491
 117492
 117493
 117494
 117495
 117496
 117497
 117498
 117499
 117500
 117501
 117502
 117503
 117504
 117505
 117506
 117507
 117508
 117509
 117510
 117511
 117512
 117513
 117514
 117515
 117516
 117517
 117518
 117519
 117520
 117521
 117522
 117523
 117524
 117525
 117526
 117527
 117528
 117529
 117530
 117531
 117532
 117533
 117534
 117535
 117536
 117537
 117538
 117539
 117540
 117541
 117542
 117543
 117544
 117545
 117546
 117547
 117548
 117549
 117550
 117551
 117552
 117553
 117554
 117555
 117556
 117557
 117558
 117559
 117560
 117561
 117562
 117563
 117564
 117565
 117566
 117567
 117568
 117569
 117570
 117571
 117572
 117573
 117574
 117575
 117576
 117577
 117578
 117579
 117580
 117581
 117582
 117583
 117584
 117585
 117586
 117587
 117588
 117589
 117590
 117591
 117592
 117593
 117594
 117595
 117596
 117597
 117598
 117599
 117600
 117601
 117602
 117603
 117604
 117605
 117606
 117607
 117608
 117609
 117610
 117611
 117612
 117613
 117614
 117615
 117616
 117617
 117618
 117619
 117620
 117621
 117622
 117623
 117624
 117625

```

gene
HOASFTTHGVCCLAPKQVAGVVDGVDASDVDSRKRFHDCGLWASSFAKSTDTPTAMAA
ATSITANHPVTPQAQSSALITFPMNRAASIDDDKDDVEVMGELYFRMNGFIIDDE"
19922. .26827
/genes="L354.8"
/notes="L4407.7; predicted using Glimmer, Testcode and
CodonUsage; Glimmer predicts 2nd ATG at 22877 is used as
start codon"
19922. .26827
/genes="L354.8"
/notes="Blastp similarity to hypothetical proteins from
several organisms"
/codon_start=1
/product="L354.8"
/protein_id="AAF70126.1"
/ab_vref="GI.780140q"

```

[illegible]

```
Query Match      28.0%; Score 114.8; DB 29; Length 38587;
Best Local Similarity 57.1%; Pred. No. 1.7e-14;
Matches 228; Conservative 0; Mismatches 168; Indels 3; Gaps 1;
```

QY	11	cggggtcaggggcagaacctgccctcaacattgcagagaagggtccccatctctgt	70
DB	30258	CGGCGCAGTGGGCAGAATCTCCCTTGAACTGCCGGAAGGGATTAAACGTTCGGT	30317

QY 71 gtacaacaggacacacctcaaggtggacagac --- cgtgcagcgtgccaaaggcagaagg 127

Db	30318	CTTCAACCGCACCTACGCGAAGACGACGTCGTTTCTCAAGGAGCATGAGACGCGAGAAATT	30317
QY	128	aaaccttcacggtctacggttccatgaaccccgctccctttgtgaagttccattcagaagcc	187

30378 TGCGGCCAACCTGAATGGATACGAGACCATGAAGAGTTTCGCTGCGTCCCTCAAGAAGCC 30437 Db
188 acgaatgatgatcatgctggtcaagaccagcgcgcagttgacagagccatcgccacgct 247 Ov

Db 30438 GCGCGCGGGTTCATTCTCGTCCAGCGCGCGCTACTGACTCTACATCGACGCT 30497

		C	T	G	A	A	G	T	G	T	T	C	G	A	G	A	C	G	S	G	G	A	C	A	T	C	A	T	A	T	C	G	A	C	A	C	T	T	C	A	A	G	G	A
Qy	248	cgcagctcatttgagcaggcgactgcatccatgatgggggaacagagtgtcacgaa	307																																									
Db	30498	CAAGGAAGTGTTCCGAGAACGGCGACATCATTAATCGCACACTTGTAACTGC	30557																																									

Qy	308	cacgagaggaggagaagccatcggaggcgcgccctnctgtatcttggcatgggtgtg	367
Db	30558	CCAGCACAACGCGCGGGCTCAGTTTGGAGAGCAGGTCTCGCTTCCTCGCATGGGCAT	30617

Qy 368 ctctgaggaaaggggtgcccgcaacggccgcctcctt 406
||||| | ||||| | ||||| | |||||
db 30618 CTCTGCTGTGGAGCAGAGGTGCGGCAGGGGCCGGCCCTT 30656

RESULT 14

AC073566/c	AC073566	42000 bp	DNA	HTG	24-JUN-2000
LOCUS	Leishmania major chromosome 35 clone L5024 strain Friedlin, ***				
DEFINITION	SEQUENCING IN PROGRESS ***, 10 unordered pieces.				
ACCESSION	AC073566				
VERSION	AC073566.1	GI:8698748			

KEYWORDS HTG; HTGS_PHASE1.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 42000)
Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P., Ivens,A., Nguyen,D., Munden,H. and Stuart,K.
Direct Submission
Submitted (24-JUN-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1990: contig of 1990 bp in length
* 1991 2023: gap of unknown length
* 2024 3664: contig of 1641 bp in length
* 3665 3697: gap of unknown length
* 3698 6805: contig of 3108 bp in length
* 6806 6838: gap of unknown length
* 6839 10632: contig of 3794 bp in length
* 10633 10664: gap of unknown length
* 10665 15449: contig of 4785 bp in length
* 15450 15481: gap of unknown length
* 15482 19168: contig of 3687 bp in length
* 19169 19200: gap of unknown length
* 19201 22797: contig of 3597 bp in length
* 22798 22829: gap of unknown length
* 22830 27115: contig of 4286 bp in length
* 27116 27147: gap of unknown length
* 27148 32918: contig of 5771 bp in length
* 32919 32950: gap of unknown length
* 32951 42000: contig of 9050 bp in length.
FEATURES Location/Qualifiers
source
1. .42000
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="L5024"
BASE COUNT 8701 a 11999 c 11887 g 9113 t 300 others
ORIGIN
Query Match 26.9%; Score 110.4; DB 57; Length 42000;
Best Local Similarity 56.4%; Pred. No. 1.3e-13;
Matches 226; Conservative 0; Mismatches 172; Indels 3; Gaps 1;
Qy 9 tccgcggtcatggggcagacccttgcctcaacattgcagagaaaggttcccatctct 68
Db 7589 TCGCGCTCACTGGCGCGAATCTCGCCCTGACATCGCCGACGAGGATTTAAAGTTGCC 7530
Qy 69 gtgtacaacaggacaacctccaaggtggacgag---accgtgcagctgccaagcgagaa 125
Db 7529 GTCTTCAACCGCCTACCGGAGACGACGCTGTTCTCAAGGACCATGAGCGGAGAA 7470
Qy 126 ggaaccttccgtctacagcttccatgaccgcgctctcttggaaagtcattcagaag 185
Db 7469 TTGCGCGCCCAACCTGAATGGATACGAGACCATGAAGGAGTTGCGTCTCGCTCCCTCAAGAAG 7410
Qy 186 ccacgggtgtgatctgctgtcaaggccgcgccgagtgaccagaccattcgcagc 245
Db 7409 CCGCGCGCGCTTCAATCTCGTCCAGCGCGCGCGCTACTGACTCTACAACTGACGAG 7350
Qy 246 ctgcagctcaacttgagcaggcgactgcatcatcgatgggggaacagtggtacgag 305
Db 7349 CTCAAGSAGTGTTCGAGAACGCGACATCATATAATCGACACTGTTGGTAATGCGAACTTCAAG 7290

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:40:13 ; Search time 4075.18 Seconds
(without alignments)
321.525 Million cell updates/sec

Title: US-09-300-482-619
Perfect score: 300

Sequence: 1 gattattgacacaccgat.....ttcagaaatgtctgtgtgg 300

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_bal:*

2: gb_ba2:*

3: gb_om:*

4: gb_ov:*

5: gb_pat:*

6: gb_ph:*

7: gb_pl1:*

8: gb_pl2:*

9: gb_pr1:*

10: gb_pr2:*

11: gb_pr3:*

12: em_fun:*

13: em_hum1:*

14: em_hum2:*

15: em_in:*

16: em_om:*

17: em_or:*

18: em_ov:*

19: em_pat:*

20: em_ph:*

21: em_pl:*

22: em_ro:*

23: em_sts:*

24: em_sy:*

25: em_un:*

26: em_vi:*

27: gb_htg1:*

28: gb_htg2:*

29: gb_in1:*

30: gb_in2:*

31: em_bal:*

32: em_ba2:*

33: em_hum3:*

34: em_hum4:*

35: gb_pr4:*

36: gb_htg3:*

37: gb_htg4:*

38: gb_htg5:*

39: gb_htg6:*

40: gb_htg7:*

41: em_htg1:*

42: em_htg2:*

43: em_htg3:*

44: em_hum5:*

45: gb_pl3:*

46: gb_pr5:*

47: gb_htg8:*

48: gb_htg9:*

49: gb_htg10:*

50: gb_htg11:*

51: gb_htg12:*

52: gb_htg13:*

53: gb_htg14:*

54: gb_in3:*

55: gb_htg15:*

56: gb_htg16:*

57: gb_htg17:*

58: em_htg4:*

59: em_htg5:*

60: em_htg6:*

61: em_htg7:*

62: em_hum6:*

63: gb_htg18:*

64: gb_htg19:*

65: gb_ba3:*

66: em_htg8:*

67: em_htg9:*

68: em_htg10:*

69: em_htg11:*

70: em_htg12:*

71: em_htg13:*

72: em_htg14:*

73: em_htg15:*

74: em_htg16:*

75: em_htg17:*

76: em_htg18:*

77: em_htg19:*

78: em_htg20:*

79: em_htg21:*

80: em_htg22:*

81: em_htg23:*

82: gb_pr6:*

83: gb_pr7:*

84: gb_htg20:*

85: gb_htg21:*

86: gb_htg22:*

87: gb_htg23:*

88: gb_ro:*

89: gb_sts1:*

90: gb_sts2:*

91: gb_sy:*

92: gb_un:*

93: gb_vil:*

94: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	% Match	Length	DB	ID	Description
1	199.2	66.4	2142	45	SOJ00265		AJ000265 Spinacia
2	79.2	26.4	4176	7	AF120494		AF120494 Arabidops
3	79.2	26.4	125803	8	ATF22K18		AL035356 Arabidops
c	4	79.2	26.4	198402	8	ATCHRIV61	AL161561 Arabidops
5	64	21.3	2792	65	SYCCVSEPGI		D13777 Synechocyst
6	64	21.3	128598	2	D90912		D90912 Synechocyst
c	7	36.2	12.1	15727	1	AE001224	AE001224 Treponema
c	8	35.6	11.9	156995	85	AL365270	AL365270 Homo sapi
c	9	35.4	11.8	101370	64	AL353604	AL353604 Homo sapi
c	10	35	11.7	11836	1	AE002356	AE002356 Chlamydia
c	11	35	11.7	162149	48	AC023104	AC023104 Homo sapi
c	12	34.6	11.5	33328	40	AC019951	AC019951 Drosophila

13 34.6 11.5 148447 36 AC010575 Drosophil
14 34.6 11.5 207585 57 AC073779 Mus muscu
15 34.6 11.5 235066 57 AC074208 Mus muscu
16 34.6 11.5 319765 29 AE003481 Drosophil
17 34.4 11.5 136570 39 AC016555 Homo sapi
18 34.4 11.5 175191 10 AC010251 Homo sapi
19 34.4 11.5 179667 50 AC025681 Homo sapi
20 33.6 11.2 155069 51 AC027133 Oryza sat
21 33.4 11.1 165921 40 AC019354 Homo sapi
22 33.4 11.1 167271 27 AC006269 Homo sapi
23 33.4 11.1 169140 52 AC036229 Homo sapi
24 33.2 11.1 146879 38 AC015095 Drosophil
25 33.2 11.1 152585 27 AC008025 Homo sapi
26 33.2 11.1 152667 52 AC027702 Homo sapi
27 33.2 11.1 158467 85 AL365446 Homo sapi
28 33.2 11.1 164107 38 AC016012 Homo sapi
29 33.2 11.1 164865 37 AC012483 Homo sapi
30 33.2 11.1 167349 39 AC016099 Homo sapi
31 33.2 11.1 168960 39 AC016100 Homo sapi
32 33.2 11.1 203924 37 AC012159 Drosophil
33 33.2 11.1 301929 29 AE003502 Drosophil
34 33 11.0 149528 87 HS367B10 Homo sapi
35 33 11.0 153140 52 AC034202 Homo sapi
36 33 11.0 157609 50 AC025466 Homo sapi
37 33 11.0 165334 28 AC008544 Homo sapi
38 33 11.0 189925 87 CNS01DX4 Homo sapi
39 33 11.0 190685 49 AC023587 Homo sapi
40 32.8 10.9 18765 1 AE000961 Homo sapi
41 32.8 10.9 19590 87 PFMA113P5 Al049182 Plasmodiu
42 32.6 10.9 137502 64 AL161737 Homo sapi
43 32.6 10.9 120180 63 AF279873 Homo sapi
44 32.6 10.9 126474 46 HS181N1 282899 Human DNA s
45 32.6 10.9 135404 63 AF228658 Homo sapi

ALIGNMENTS

RESULT 1
SOJ00265 2142 bp mRNA PLN 25-AUG-1998
LOCUS Spinacia oleracea mRNA (nuclear-encoded) for chloroplast
DEFINITION glucose-6-phosphate isomerase.
ACCESSION AJ000265
VERSION AJ000265.1 GI:3413510
KEYWORDS glucose-6-phosphate isomerase.
SOURCE Spinach.
ORGANISM Spinacia oleracea
REFERENCE Martin, W.F.
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
TITLE Euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
JOURNAL Caryophyllales; Chenopodiales; Spinacia.
REFERENCE 1 (bases 1 to 2142)
AUTHORS Direct Submission
TITLE Submitted (16-DEC-1997) Martin W.F., Spielmannstr. 7, Institut fuer
JOURNAL Genetik, TU Braunschweig, D-38023 Braunschweig, D-38023, GERMANY
REFERENCE 2 (bases 1 to 2142)
AUTHORS Nowitzki, U., Flechner, A., Kellermann, J., Hasegawa, M.,
TITLE Schnarrenberger, C. and Martin, W.
JOURNAL Eubacterial origin of nuclear genes for chloroplast and cytosolic
glucose-6-phosphate isomerase from spinach: sampling eubacterial
gene diversity in eukaryotic chromosomes through symbiosis
Gene 214 (1-2), 205-213 (1998)
FEATURES
89322268 Location/Qualifiers
1..2142
/organism="Spinacia oleracea"
/db_xref="taxon:3562"
/note="nuclear gene for chloroplast enzyme"
transit_peptide 84..263
/gene="GPIP"
/product="nuclear encoded chloroplast GPI"
84..1940
CDS

/gene="GPIP"
/EC_number="5.3.1.9"
/function="glycolysis, gluconeogenesis"
/note="nuclear encoded chloroplast GPI"
/codon_start=1
/product="glucose-6-phosphate isomerase"
/protein_id="CAA03982.1"
/db_xref="GI:3413511"
/db_xref="SPTREMBL:O82058"
/translation="MASSLSNLYSTSLKPKISHLPTITKNPISPKSLSPKPISSVA
RDTADLSTSSSSNNLPSLQKKADSLKEDPALMARVEMLYOHKDLGLVDVS
RIGFDEFVKEMEPERFAKFMIEGIEKGAIANDEGRMWHYWLRLSSLAPTFLKN
QIDVLDKRWQFANDVI SGKT KAPTGERFTHLSVIGGSGALGPOFAEALAPNDGLV
KTRFDNDTPAGIDHQAOLGPELATITLVVIVISKGGTPETRNGLLEVKAFADGALV
FKAGVATQENSLDNTARTGWDREPFMDWGGRTSEMSAVGLLPAALQGDIDKE
MLAGALMDEATKIPVLSNPAALLANMYASDGVSKDMVLPYKDSLLLSRYLQ
QIVMSLEKPEFLDKNKNVQGLTVYVGNKSGSDOHAHQIOLRGVHNFATFIEVLDR
PFGHSELEPGVTGCDYLFGLMGTQVTSALYANNRESISVTQVETVPSVGMVALYER
AVGLYASLVNINAYHOPGVGEAKAAAEVLALQKRVLAVLNEASCKDPVEPLTIEEVA
DHCHCPDDITMIYKIIAHMAANDRVILAEAGDCGSPRSIKAFLGECNVDELVA"
84..1940
/gene="GPIP"
mat_peptide 264..1937
/gene="GPIP"
/EC_number="5.3.1.9"
/note="nuclear encoded chloroplast GPI"
/function="glycolysis, gluconeogenesis"
/product="glucose-6-phosphate isomerase"
BASE COUNT 593 a 467 c 490 g 592 t
ORIGIN
Query Match 66.4%; Score 199.2; DB 45; Length 2142;
Best Local Similarity 79.0%; Pred. No. 8.8e-51;
Matches 237; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 1 gattattgacacacgcgactctctgctgggattgatacaaaattgctcaactagacgctg 60
Db 745 GATTATTGATAATACAGACCCAGCTGGAATCGATCATCAAAATGACAACTTGGCTCG 804
QY 61 aactggcaactactctgttaattgcttaagcagcagcagcagcagcagcagcagcagc 120
Db 805 AGCTAGCTACACACTTGTATGTTTTCARAGAGTGGCGGTACACCAACAAAGAA 864
QY 121 atggtctactagaagtacagaagccttcagagatgcggggctgcgaattctcgaacagg 180
Db 865 ATGTTTATTGGAAGTACAAAAGCGCTTCGGTGTGATTAGTAGTATTCGCAAAACAGG 924
QY 181 gtgtgcaactactcaagaataattctctgttgataacactgctagatagaggaggt 240
Db 925 GTGTGCTATTAACACAGAAACTCATTTGTAGATATATCTGCACGAAATGAAGGCTGA 984
QY 241 tagctcggttctctatgtttgattgggtggtggtggtggtggtggtggtggtggtggtggt 300
Db 985 TAGATAGATTTCCCATGTTTCGATTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1044
RESULT 2
LOCUS AF120494 4176 bp DNA PLN 12-JAN-2000
DEFINITION Arabidopsis thaliana phosphoglucose isomerase precursor (PGI) gene,
complete cds; nuclear gene for chloroplast product.
ACCESSION AF120494
VERSION AF120494.1 GI:6690394
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
MAGNOLIOPHYTES; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 4176)
YU.T.S., Yue.W.L., Wang,S.M. and Chen,J.
Mutation of Arabidopsis chloroplast phosphoglucose isomerase
TITLE

affects starch synthesis and floral initiation

JOURNAL
Unpublished
2 (bases 1 to 4176)
Yu, T.S., Lue, W.L., Wang, S.M. and Chen, J.
TITLE
Direct Submission
JOURNAL
Submitted (15-JAN-1999) Institute of Molecular Biology, Academia
Sinica, Taipei, Taiwan 11529, Republic of China
FEATURES
Location/Qualifiers
1. 4176
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="4"
/map="67.3 cm"
Join(338..856,1067..1186,1384..1482,1566..1649,
1932..2076,2146..2210,2365..2433,2522..2629,2891..2943,
3034..3172,3259..3415,3507..3585,3675..3741,3827..3961)
/gene="PGI"
/product="phosphoglucose isomerase precursor"
338..3961
/gene="PGI"
Join(338..856,1067..1186,1384..1482,1566..1649,1932..2076,
2146..2210,2365..2433,2522..2629,2891..2943,3034..3172,
3259..3415,3507..3585,3675..3741,3827..3961)
/gene="PGI"
/codon_start=1
/product="phosphoglucose isomerase precursor"
/protein_id="AA24124.1"
/db_xref="GI:6690395"
/translation="MASLGLVSSPSLKPKNHFKGIAGLSRPLFPVLPQHS
TLTLSSARSARDISUADSKELLKDPDALWRYLDWFYQKELGLYLDISRVGTDE
FVAMERFQAAPKAMEDLEKSGFIANDPDEGRVMGHVWLRNSKLAPKPLKLTENTLD
SICAFSDIIISGKIPSPPEGRFTQILSVGIGGALGQSLRLGPDNPPLKIRFID
NTPAGIDHQAIGLPELSTVIVVLSKGGTPETRNGLEVKQAFREAGLNFAGQV
AITQENSLONTARIEGLARFPMYDMVGRSTLSMSVCLLPAHLOGINVREMLTGA
LMDEARTTSIKNPAALLAMCWYASNGVSKDMVLPYKDSLLFLSRYLQQLVMS
LGREFLDGNTVNOGLTVYNGKSTDOHYTQQLDKDGVHNFATFIEVIRYRPPGHDW
ELEPGVTCGYLFGMLQGRSALYANGRESISVTOEVTPTSGVAILALYERAVGLYA
SYVNINAHOPGVEAGKAAAEVLALOKRVLNLEATKDPVEPLTLEETADRCAP
EELEMIYKIIAHSANDRVLIAGNCGSPRSIKVYLGECNVDDLYA"
transit_peptide 338..556
/gene="PGI"
polyA_site 4109
polyA_site 4176
BASE COUNT 1153 a 829 c 859 g 1335 t
ORIGIN

Query Match 26.4%; Score 79.2; DB 7; Length 4176;
Best Local Similarity 77.4%; Pred. No. 1e-13; 28; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 177 cagggtgttcaattactcaagaaaattctctgttgataaacactgctagaatagaggga 236
|||||
Db 1929 CAGGGTGTTCATAACCCAGAGAAGACTCATCTACTGATAATACGCAAGATTGAAGT 1988
|||||
QY 237 tggtagctgggttctactatgttgattgggttggttgtagacttcgaaatgctgct 296
|||||
Db 1989 TGCTAGCTAGATTCTCTATGTACGACTGGGTGGGGAAGAACAATCAATAATGCTGCA 2048
|||||
QY 297 gtgg 300
|||
Db 2049 GTTG 2052

RESULT 3
ATF22K18 125803 bp DNA PLN 03-FEB-1999
LOCUS
DEFINITION
Arabidopsis thaliana DNA chromosome 4, BAC clone F22K18 (ESSAI
project).
ACCESSION
AL035356
VERSION
AL035356.1 GI:4220510
KEYWORDS
thale cress.
SOURCE

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
Brassicaceae; Arabidopsids.
1 (bases 1 to 125803)
Bevan, M., Wedler, H., Wedler, K., Wambutt, E., Hoheisel, J.,
Mewes, H.W., Mayer, H.F.X. and Schueller, C.
Unpublished

2 (bases 1 to 125803)
EU Arabidopsis sequencing project.
Submitted (03-FEB-1999) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 125803
Location/Qualifiers
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
171..2980
/gene="F22K18.10"
complement(join(171..601,689..883,983..1159,1250..1469,
1556..2980))
/gene="F22K18.10"
complement(171..601)
/gene="F22K18.10"
/number=1
complement(join(171..601,689..883,983..1159,1250..1469,
1556..2980))
/gene="F22K18.10"
/note="similarity to DNA polymerase III gamma subunit -
Aquifex aeolicus, PIR2:A70460
Contains ATP/GTP-binding site motif A (P-loop) [PRTGTGT]"
/codon_start=1
/product="DNA polymerase III like protein"
/protein_id="CAA22984.1"
/db_xref="GI:4220511"
/translation="MGTRRSHSDVPIRITLVALRRVRLRDPCTTSMKSFASLLDNV
KWETGNSNGISLQFVHADDAACKAAADAPVGLIPGYSYIMELESGLDLKLSKVI
NVEGDACSRSSRSCDLSVGRDLACNAPSIHVEAGSGGRYRTHYTKLASSVGE
YGRSLGSPMNSTHYSYGDVDVDEDSQNRCCGTYCWSRTPRPNQSSDVEEYPL
LPGNGESDVVTPSHVEVLSLSQKFRPKSFDLVGQVGVKCLLTILGRITTSYV
LPHGPRGTSTSKIFAAALNCISQAHSRPGCLCECKSYFSGRGRDWNEDSGKL
NRPSYLSLTKASLPVSSRFKVIIDECQLQCTGTGLLNSLDFNSOHSFVLVT
SELEKLPNVLRSQKHYFVKCDADISTKLAKICIEGIDFGQAVDFATASDGL
RDAETMLDLQLSLGKRTTSLAYKLIGVSDDELLDLIAMSDDTSNTVIRARELMR
SKIDPMOLISOLANVIMDIAGNSQESSATRLRLTRHTESEEMKRLNALKILSDA
EKHLRAKNOTTLTVALLOLNTDSSSFATDENGRIKNDVELSTSSGCPGVK
SDAEKQGRNENVEVSKVTIDCCSDSLKRLFKRLTSLVDKGVIAELFVY
TPQHVAREKSKLIDAFQSVLGCNVEIQMLVISACPPKAKAAASFFGLFSCS
RMLKLSYLTTRTSDSCAKEPAYTNLSRCQGNLRSVRSANASRRMSCSSDQG
DATSAMCTPHIPGGEK"
complement(602..688)
/number=1
complement(689..883)
/gene="F22K18.10"
/number=2
complement(884..982)
/number=2
complement(983..1159)
/gene="F22K18.10"
/number=3
complement(1160..1249)
/number=3
complement(1250..1469)
/gene="F22K18.10"

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
Brassicaceae; Arabidopsids.
1 (bases 4709 to 5139)
Robben, J., Grymonprez, B., Voickaert, G., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
Unpublished
2 (bases 16658 to 125116)
Zimmermann, W., Gruenisen, A., Wambutt, R., Kalicki, J., Wohldmann, P.,
Smith, A., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.
Unpublished
3 (bases 123107 to 198402)
Wedler, H., Wedler, E., Wambutt, R., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
Unpublished
4 (bases 1 to 20862)
Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W.,
Lemcke, K. and Mayer, K.F.X.
Unpublished
5 (bases 1 to 198402)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk

REFERENCE
AUTHORSJOURNAL
AUTHORSJOURNAL
AUTHORSREFERENCE
AUTHORSJOURNAL
AUTHORSREFERENCE
AUTHORS

COMMENT

Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV60 at the 5' end and an
overlap with ATCHRIV62 at the 3' end.

FEATURES
source

1..198402
Location/Qualifiers

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

7612..7808

/gene="AT4g24160"

/number=1

7612..9918

/gene="AT4g24160"

join(7612..7808,8031..8267,8480..8526,8657..8775,
8876..9036,9206..9315,9402..9501,9574..9697,9790..9918)

/gene="AT4g24160"

/note="similarity to CGI-58 protein - Homo sapiens,
PID:94929585"

Contains Prokaryotic membrane lipoprotein lipid attachment
site AA376-386

contains EST gb:AI099669, T21097, AA394561, T76500,
T42730"

/codon_start=1

/product="putative protein"

/protein_id="CAB79326.1"

/db_xref="GI:7269266"

/translation="MAEISKTKVSGSSSTASVSSAASAAATNAKSRKILWPNLS
RWIPSTDIYIAAEKRLILKTPYVOEYSGSPGSKIRFRSTNSREYINVT
FDKAGAPTLNVGICGASGSGFFERNDALASRPVTAIDLGWGGSGDFTCRSTE
ETEAWFIDSEFNKRAQNLNFIILHGSFGYVAAKYLKHPHGVHQLILVGSAGFSA
EADAKSEWLTFRATWKGAVLNHLWESNFTPKVLVGLGFGPLGVNRYTARFGAHS
ETGTEEAALLDYVYHTLAARASGELCIKYTFSPGAFARPLQASAEWKPVTPE
IYGMNWNQYGAWEAKSRMKVPCEIIRVFGGHHVFIDNPFGHSAVLACRKFISQ
DSSHDQQLLDGLRLV"

7809..8030

/gene="AT4g24160"

/number=1

8031..8267

/gene="AT4g24160"

/number=2

8268..8479

intron

exon

intron

/gene="AT4g24160"
/number=2
8480..8526
/gene="AT4g24160"
/number=3
8527..8656
/gene="AT4g24160"
/number=3
8657..8775
/gene="AT4g24160"
/number=4
8776..8875
/gene="AT4g24160"
/number=4
8876..9036
/gene="AT4g24160"
/number=5
9037..9205
/gene="AT4g24160"
/number=5
9206..9315
/gene="AT4g24160"
/number=6
9316..9401
/gene="AT4g24160"
/number=6
9402..9501
/gene="AT4g24160"
/number=7
9502..9573
/gene="AT4g24160"
/number=7
9574..9697
/gene="AT4g24160"
/number=8
9698..9789
/gene="AT4g24160"
/number=8
9790..9918
/gene="AT4g24160"
/number=9
10914..16518
/gene="AT4g24170"
10914..11070
/gene="AT4g24170"
/number=1
join(10914..11070,11153..11242,11317..11417,11501..11620,
11715..11796,11890..11957,12026..12213,12305..14416,
14692..14793,14875..15081,15183..15231,15323..15374,
15484..15656,16085..16198,16342..16518)
/gene="AT4g24170"
/note="similarity to kinesin heavy chain
-Strongylocentrotus purpuratus, PID:gl0270
Contains ATP/GTP-binding site motif A (P-loop)
AA89-96,kinesin motor domain signature and profile
AA225-236"
/codon_start=1
/product="putative protein"
/protein_id="CAB79327.1"
/db_xref="GI:7269267"
/translation="MGGEKILSVRVPLNEKEKTRNDRCDWECINDTITICKFHNH
PKSSYFDKVFGECPKQYDDGAEVLCVLSGINSIFAYGOTSSTGKTYTMSGI
TEFAMDDIFAYIDKHQERKPTLKFSAMEIYNENAVRDLLCEDSTPLRLDDPGRSTV
VEKREETLRDRSHLELLSICETQKIGETSLNEISSRSHQILRLTIFSSQOQFSPE
SSATLAASVCFVLAGSERASQTLASGRLEKEGHNRSLTLGTWIKLSKGNKGI
PYRDKTRILIONSLGNARTALICTMSPARSHLEQSRNLTLLFATCAKEVTYNAQVNL
VSEKALVKOLRELARWENELNGLPASSTDFYALMKQKEELIAKMEEQIHEL
KWQRDVAQSRVNLKTAERSSSSMSDRRRRIYDSTDFDEPMLNNGKSNLYS
PDGFLDDTTPQFNGHNLHKWEEMAQSTTQEPEDACKVEKIVNNGEAEVIO
DSLDDIVERKEYESQKDDADSSIKNIDMELSLYTKPEADGYSVKLLDVOET
EQSVKQKSPKPEMEQYLSRDMSEQVTKSLPEEQCYEYCAVDKLEAODVFLNK
LEESQQTQSVKEKEDTKNLSSKKEDLKNLSMDQSEQLYKSPPEDEKVCVEYEGSDK
DDNTYEALKKKKEMQKTIIEYFMSIQSAEEKQSPFNIDDTLSPGEYFMRKRSRSCR

Db 1711 ATGCTCTGCGGAAACGAAGCAGTATTGTAAGCCAGCGCTTCATTTGCCGATATG 1770

QY 181 gtattgaattactcaagaaattctctgttgataacactgctagaatagaggggt 240

Db 1771 CGTGGCGGTGACCATGCGCGCAGCATTAATTTCCAGCAGCGCCAGCAGCAATGGT 1830

QY 241 tagctcggttctctactatgttgggtgtggtgtagaacttcagaatgctctgctgtg 300

Db 1831 TACAAGCTTTCCCATGCGAGTATGGGTGGTGGGAGAACCTCGAGTATGCGCGGTG 1890

RESULT 6

D90912 128598 bp DNA BCT 07-FEB-1999

LOCUS Synecocystis sp. PCC6803 complete genome, 14/27, 1719644-1848241.

DEFINITION D90912 AB001339

ACCESSION D90912.1 GI:1653228

VERSION 3-chlorobenzate-3,4-dioxygenase; 30S ribosomal protein S1; 30S

KEYWORDS ribosomal protein S2; 50S ribosomal protein L34; 50S ribosomal

protein L9; ABC transporter; COB protein; DNA primase; GDP-mannose

pyrophosphorylase; GTP cyclohydrolase II; H+/Ca2+ exchanger; PetG

subunit of the cytochrome b6f complex;

UDP-N-acetylmutamoylalanyl-D-glutamyl-2,

6-diaminopimelate-D-alanyl-D-alanine ligase; acetyltransferase;

anthranilate synthase component I; aspartate carbamoyltransferase;

carbonic anhydrase; catabolite gene activator protein; cyclase;

cytochrome CytM; cytochrome c oxidase folding protein; elongation

factor TS; extragenic suppressor SubB; fatty acid desaturase;

ferredoxin; glucose-6-phosphate isomerase; heme oxygenase;

hemolysin; hydrogenase expression/formation protein HycP; inner

membrane protein; melibiose carrier protein; membrane bound protein

LYR; ornithine acetyltransferase; oxygen independent

photophosphorylation III oxidase; polyA polymerase; regulatory components

of sensory transduction system; renin-binding protein; sensory

transduction histidine kinase; serine acetyltransferase; tRNA-Arg;

tRNA-Ile; tRNA-Ser; transposase; tyrosine protein cthy28kD; AT103.

Synecocystis sp. (strain:PCC6803) DNA.

Synecocystis sp.

ORGANISM Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

REFERENCE 1 (sites)

AUTHORS Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,

Miyajima,N., Hirose,A., Sugita,M., Sasamoto,S., Kimura,T.,

Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,

Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,

Yamada,M., Yasuda,M. and Tabata,S.

TITLE Sequence analysis of the genome of the unicellular cyanobacterium

Synecocystis sp. strain PCC6803. II. Sequence determination of the

entire genome and assignment of potential protein-coding regions

DNA Res. 3 (3), 109-136 (1996)

JOURNAL 97061201

MEDLINE 2 (bases 1 to 128598)

REFERENCE Tabata,S.

AUTHORS Direct Submission

TITLE Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshi

JOURNAL Tabata, Kazusa DNA Research Institute, Laboratory of Gene Structure

2; 1523-3, Yanauchino, Kisarazu, Chiba 292, Japan

(E-mail:tabata@kazusa.or.jp. Tel:+81-438-52-3933,

Fax:+81-438-52-3934)

COMMENT Potential protein coding regions were assigned on the basis of

similarity search of the ORFs and Genemark analysis.

FEATURES Location/Qualifiers

source 1..128598

/organism="Synecocystis sp."

/strain="PCC6803"

/db_xref="taxon:1143"

complement(1..933)

/note="ORF_ID:s111263"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAA18144.1"

/db_xref="GI:1653229"

CDS

/translation="MTARLPPTAVLSIGAAALATMGKLGAYAITGVSGLLSDALEST

VNLSAIVAFWALSLAATPADSEHPGHKAIEYFSSGLEGAIFVAALGIGYSVERL

LSPRLDOALGIALAATAALNGTAVILWRAGKRLNSLDRSQHLMQDVTWTSVG

VVAVAILFTVGNWLDPLIALGVGNVLTGTHLLRETISLMDQSLPAPLOAITS

CFLEDEQGVRFHLLTROAGSQSIFSHVLVPGHWTVQRGHDLCEAIEATAERTIG

SRVTHLEPLEDPKSMOHPDEFPPSAPLNDRKPN"

complement(997..1482)

/note="ORF_ID:s111262"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAA18145.1"

/db_xref="GI:1653230"

/translation="MLPLPIAKNGKGFIRALENDGALAVYAPLEGYEGYORRLRAN

YASISLSARGLGDVEAYLMQVHGVRPAHLGKNKNTAAGEAGVPIYFAQPIAGYQLENL

GYASKGLVILWLEGYILISQTEIQDLSLTKRVPKLVLEMGGDRVFRWQPLDLCQA

A"

complement(1707..2363)

/gene="tsf"

complement(1707..2363)

/gene="tsf"

/note="ORF_ID:s111261"

/codon_start=1

/transl_table=11

/product="elongation factor TS"

/protein_id="BAA18146.1"

/db_xref="GI:1653231"

/translation="MAEITQALVKELREKTGAGMMDCKKALKENEGDLEKSTEWLRQK

GIASAKKSGRTAAEGLVSIYHFGRIQVIVENCETDFVARGDRFDLVNDVAMQI

AQPNYIVSVADIPOEMVAKEKEITEMGRDDLGGKPKANIKKIVQGRDKRLKLSLL

"

complement(2517..3326)

/gene="rps2"

complement(2517..3326)

/gene="rps2"

/note="ORF_ID:s111260"

/codon_start=1

/transl_table=11

/product="30S ribosomal protein S2"

/protein_id="BAA18147.1"

/db_xref="GI:1653232"

/translation="MPVVSADLLESQVHFGHQTTRWNPMDQYIYARNGVHIDLV

QTAQMEDAYEYVRSTNGKFKFVGTGKQAAGTISOEACRGAHVYVNOGLGMLT

NNETIRKVRDLKELEALEASGIDRRGKKEGSMRLRELKLOKYLGLKNNRWKLPDV

VIVYQRREHNAIHECQKLGPIIAMLDTNCDPDVDVPIPAINDAITSRLIVGKLA

DAIYEGRGQPDVSDYDEFEDEGLDGNLEVEAAEEVEAAEVAATPA"

complement(3551..3623)

/product="tRNA-Ile"

3809..4927

/note="ORF_ID:s1r1336"

/codon_start=1

/transl_table=11

/product="H+/Ca2+ exchanger"

/protein_id="BAA18148.1"

/db_xref="GI:1653233"

/translation="MSTKSKIFLVLLVFCPLSFAAHLWGWTTFVFLAGLAIPLAA

FMGTATEIAVIGPNAGGLNATFGNATELILAYIALKEGLLEVYKATLTGSIIGNL

LIVMGFAVFLGGLRYKEONFOPLAARLNASTMNLGVVAILPTALOYSTGVETVLQ

NLSVAVLVGLVGLLSVFSMGTTHAVLYDVGVAENMEMFELGDESEPPPEEKP

NLWLWTGVLVYTLGVAVESLLVGSLEVAESIGLTALEFTGVIVLPPIIGNAAEHATA

VTVAKDKMDSMSYVMGSSQLQIAFFVAVLVIVGWAIGQPMDLNFPFELVAVLV

LIVNISDGTGTSNWLLEGLILLATYAIVALAFFHPTLV"

complement(4954..5535)

/note="ORF_ID:s111258"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="BAA18149.1"

/db_xref="GI:1653234"

/translation="MIKNDIWIARAQAQGMIEPPEPKLMQIEGQPAISFGLSSFGYD

IRLSAADFRIFRHPGTVPDKNPNHLEATOLHRDSDGSYFLLPAHSGYGLGVAVER

LAIPANVTLCIGKSTVARAGIANTLTPGAGCGHLTLEFSSSSADCRITANEGIV

CDS	QLLFFEGEDCDISYETRRGKYQNPQSVTLPRV" complement(5788..6282) /note="ORF_ID:s111257" /codon_start=1 /transl_table=11 /product="transposase" /protein_id="BAA18150.1" /db_xref="GI:1653235" /translation="MLREILKLYGSAIYVYIDSGFEAIQACIYAWSKKGKVVYDGRQ GKGVRLNLAGRRKKGLDIAPWFTGSLNAGFEGWLKLVLLPSLDIPSLIMDNA PIHKTAIKELAKEAGHEVFLPKYSPDLNDIEHDFSAKLRMYAPIDTSLSDEIIRS YCGV" complement(6276..6635) /note="ORF_ID:s111256" /codon_start=1 /transl_table=11 /product="transposase" /protein_id="BAA18151.1" /db_xref="GI:1653236" /translation="MAYSLDLQRVAVYIEAGKIIEASKIYKIGKASYRWLNRYDL SPTKVERRHKLDEALKDVEENPDARLIDRAKKFGEVRPSAVIYALKMKKINKKKKN YVIEKGTGRNELSTIEC" 7054..7656 /note="ORF_ID:s1r1338" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA18152.1" /db_xref="GI:1653237" /translation="MSNORDYTLIDKSGMSITEPKFKSRWELVQESTLALARKCD OLDAGITVTFSGKFRYDNNVASKVEQIFQENEPVGGTNLTAVALQDALNLFQRRK SNQAPTGETIIVIDGPNDRSRVFELIIQASKCLDADESLAIFSMQIGNDPSATKFL QALDQLMEVGAKEFDIVDTVFDEMEDLTLEVLNNAED" 7640..8098 /note="ORF_ID:s1r1339" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA18153.1" /db_xref="GI:1653238" /translation="MLRIETRYTFIMNNSDISIDLADLRQKYIEEPETLSSNPAIAS TSNKRQKPTDIILNELKQGFVNFSEKTPKQATFKOTSESLNPELKKLIDHOSQ QROATIAOAKOQWLEQLDPLSGEIGWIFADLAKHYFSTLAAAILNNST" 8127..8663 /note="ORF_ID:s1r1340" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA18154.1" /db_xref="GI:1653239" /translation="MIDIVKADLNSAVHAEAMIOLMNEYACDPMGGGEELPNYVKANL PAELAKRSAHILAFVDSKPAAGLLVLCLEGSTFACKPLLNHIDVIVSLPYRGKGLSK LMLQKAEATLDLGGCKUTLEVEGNHVAQSAYSFGFNGYELDPQMGKALFWQKKLP DISTLDSHVSSIMIKODS" complement(9369..10385) /note="ORF_ID:s111255" /codon_start=1 /transl_table=11 /product="transposase" /protein_id="BAA18155.1" /db_xref="GI:1653240" /translation="MISNFGHIVKTYLSNFPKDDYPLVDTFKFVSIWGLMLDOSQTS MRSMKRLNRCGETVIDSFASKKRDVGFREIFSLKKELSKRKEIKOGELEIFP LDSTIVTSKLMWNLGHQVVFSGINLSTPGGVIHFQGHDKNYGNETIETP ENGVAWMDGFCDLQRIKLEKNKYHVLRIKNIKLEKLANDNYMGTGKNKIESR VVFTHDSEFRVLNLPDIESIEGVSDEKTAELYKKRWQIEFLKFLKMLKLNRL IAKNENAIQIQTICLAIYLLIKLVIFKEAGTTLMDLRLYLAQFMCEKISYVHWLRE
CDS	complement(21.3%; Score 64; DB 2; Length 128598; Best Local Similarity 54.2%; Pred. No. 7.5e-09; Matches 130; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
CDS	Qy 61 aactggcaactactcttgaattgtcaatttcttaagagcgaggacacacctgaaccgcga 120
CDS	21.3%; Score 64; DB 2; Length 128598; Best Local Similarity 54.2%; Pred. No. 7.5e-09; Matches 130; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC65461.1"
/db_xref="GI:3322762"
/translation="MSGHSKWATIKHAKGAADAKRGOLFTKFIKEISVAARMAGGQPO
ANPRILAILKARANMPKNDIERAIKKGTCGELSSGVSEELVYEGYAGGVAVLVEVL
TNKRAANVNRLFSRNGNLGAGSVSYNFKGVIEDSEQVDEALMELALEAG
AEDIQNAGCVLTVTVPCFTFVLESQAQWESLSAGISNVPDYLALDEETARKVL
KNIDLEEDVOAVYSNADIPSELV"
complement(1655..3262)
/gene="Tp0475"
/CDS
complement(1655..3262)
/gene="Tp0475"
/notes="similar to GB:AE000783 percent identity: 46.56;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate isomerase (gpi)"
/protein_id="AAC65462.1"
/db_xref="GI:3322763"
/translation="MNRNLDCAAYARLQAIRAPSLKTVLCPEGIERVRYCTDAG
AGLRVHYAAKTVNEILITLALADEQELVAKYDALRAGAINTGKRVLHLTRLIG
VQSSLASLCEQWDMHAFYTKYERVCAFAQVHESGLRSGAFPTDVQIGIGS
DLGPRALYLEGWAQRQAVKMRTHFISNVDPDAALVLSKPLETLTFLIVSKSGT
TLETLSNELFVHVRQAGLEPHTFQVAVTSETSPANNPOYLASVYMDDFIGGRYSS
SSVCGAVLTITAFGPQVGFHFLSGAARADRAQOEDIRRNALLDALIGVETILIGY
BHTAVLPYSQALREPAHLOQLOMESNGKSVNRFQIPITYKTGPVIFGEPGNGQHSF
YQHLQGTSVVPLQFIAFQHSQGLQDPIIRKSTGQOKLLANVAQIVAFARKEHADA
NKTSGRPSLLYAKALTPTOTIGALLAHFENKIMFQGFANWLNSPDQEGVQLGKTLA
QHILAGEVGVRLAYADLFDLAHPTC"
complement(3326..4672)
/gene="Tp0476"
/CDS
complement(3326..4672)
/gene="Tp0476"
/notes="similar to GB:L23147 SP:P38502 PID:349834 percent
identity: 43.61; identified by sequence similarity;
putative"

```

gene

CDS

```

/codon_start=1
/transl_table=11
/product="acetate kinase (ack)"
/protein_id="AAC65463.1"
/db_xref="GI:3322764"
/translation="MILLTLCGSSSVKYQVYNTERAVIAGVVERVTOAGSVITHE
VHGENHVRSPCTHTEAVQLIISTLDHAGVSDMSLINAUGHRVHLHAADREVKS
VWVTPDLEFRSVQDLGPHNPANIRGTEAAQAVIPSPHCAIMDTAHQHTMPEASF
MYAVPQVVEYAVRVRGHTGTSFLYAKRASVILKRAEDFNIIAHIGNASMCVV
KNGVSDTSMGITPLBGLMTRCGDCDPALAFYIMRKTGMSVSDIDITLAKQSGLLG
ITGKYDRDRCAAMKNGDALARLAFDMEVHRIKRYFGAYVAALQKPDALVFTAGVG
EMCCDVPAAACAGLEHLGRLDARKNELARNAETEISDDSPVRLIVPTDELVN
TEDTIALMGSTYDVHTRFTYFQDPTVYNKARAAGLKRDLKKNPALATIVYPRLDT"
complement(4739..5464)
/gene="Tp0477"
/CDS
complement(4739..5464)
/gene="Tp0477"
/notes="similar to GP:1651209 percent identity: 37.04;
identified by sequence similarity; putative"

```

gene

CDS

```

/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase, putative"
/protein_id="AAC65464.1"
/db_xref="GI:3322765"
/translation="MKKHLFDARAIAAFLVSFVSDSRILKTOEILWLALSGSGTPEIF
RTWAHFERHLLQWLKRLRFWSDCVRPPTDAQSNFMTSHALLEPLEINPAVFRVVG
EAPESACAAYSQIEARLPQRGVCFCDIILGAGDGHFTASIFPHEIELWDHSGCC
VATHPDTCOKRVSTFTGHLINNAHEIYVVVTGREKQDMLASVADPHASVPLHAWTLP
KRSQWIPLLRLGLYPRKQCKPTV"
complement(5525..7072)
/gene="Tp0478"
/CDS
complement(5525..7072)
/gene="Tp0478"
/notes="similar to GB:142023 SP:P44311 PID:1004000
PID:122494 PID:1204809 percent identity: 49.38;

```

gene

CDS

```

identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase (zwf)"
/protein_id="AAC65465.1"
/db_xref="GI:3322766"
/translation="MGKISGSGTVAPHLIVFGASGLAARKLIPSLWDLFEQELLPR
TFGILGAGRTALSTESFRARAEAVTHKAVTDPHDPARLTFEQKHYFSFDTDSVA
FADFATYVTRTLDQSLHTEGNEFIVLATPPSLYETIPTQLAMHHLNRQGNFRVVEK
PEGYNLETAQHNLASLRHFOENQTYRIDHYLGKETVQNLIVTFRANPLPEPTWNRTH
IDYVEITASISGVENRGYYDQSGALRDMIONHLLLLGI IAMEAPAVSSSRLRDE
LVKVFDCLRPMGERDVMQHTVRAQVAGKIRGVAVPGYLEESGVDPRSCPTFAALKC
YTDNRWMDVPFYLRTKRPLTGVTEIVHVTLPIALFHERPCAREGNALVIRIQ
PDEGTQLKIDLKEPGAGFKTIPVSVDFQYSALYSHLPASVAYELLDCMNGDNTLHR
DDAVESAWREIDPILAAWKSNNKSPLLTPAGSWGPKAADLNGKAPRHHPSSTILLS
DDEACRL"
complement(7145..7819)
/gene="Tp0479"
/CDS
complement(7145..7819)
/gene="Tp0479"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="T. pallidum predicted coding region TP0479"
/protein_id="AAC65468.1"
/db_xref="GI:3322769"
/translation="MTLSHPTRDRSRRLVQLGLDRMGIQWHRTARGGVARCVVRVG
SLAALGMAGVAAAEVFSVKVQGAHPWKEPAPRTDVLITPTLGLAVGVIAHN
GFAPTTDLDAAGLAFMRAQFLVGNWAKIRSGKNFAFIPSLGVDILSTQDDKVGIPTNL
DLIFFTFKHGLDFTFGSGVNVPLTRNLKKGANGQDLTWSINAWKCTSLVWVKI
GPVPRI"
7931..8419
/gene="Tp0480"
7931..8419
/gene="Tp0480"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="T. pallidum predicted coding region TP0480"
/protein_id="AAC65469.1"
/db_xref="GI:3322770"
/translation="MIRALFSLRSHANTHPADLAHAALALALALPRSSLLWYLL
FVCFIRLNRLGLLSLVLFGEVFPSEDPWLDLGNWALCLPRLOQVYRALIEIPFV
GLARFYNTIAGLVAGALCYLPCYALARCNAVTAVRTVLYPKIHHTIFELVRNAPLC
KR"
8451..9884
/gene="Tp0481"
8451..9884
/gene="Tp0481"
/notes="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="T. pallidum predicted coding region TP0481"
/protein_id="AAC65470.1"
/db_xref="GI:3322771"
/translation="MSDDSTPKTPSRRIHRTGRRRTLHFFCKRVTPRSLKRLRRIH
IPADRAYCMRYLADPVPSTVRVFGTLLSRVVRPDQAIASADLKRNLNIAASIAK
ORGRVNWLSNACASVLALGLVLYLRNVIARRVYIGSGEAFGACAEAAVVDLDF
NARERLKYAVANKHHPNNWFEIESIDIHEDLLELSRGKEVSHVWVYVTPRK
TSGALPPRAKQRVRSSNPLIAKIOEKAAEAAPVSEFGAFSALKAQVDPRLLERE
VKALPTTLVHQVGAQAPKLAIRQVDFDAARAETAAAIIRAVTELDJHAKLDVSA
IKQETTDIADARRTEALATIRSHELQDDVHSTLGLAREFAAAVKADGARTARA
AAIRIDIQDGGKKFISGLCTVFLARFSHYPPYVAQMLDYVRGSGRTPSDGSFAEAE
KTAOSLTTRKRLOGVIFCLSATSLPC"
9851..10699
/gene="Tp0482"
9851..10699
/gene="Tp0482"
/notes="hypothetical protein; identified by Glimmer;

```

gene

CDS

Sat Nov 4 18:11:41 2000

```

Query Match      12.1%; Score 36.2; DB 1; Length 15727;
Best Local Similarity 51.6%; Pred. No. 2.2;
Matches 83; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 63 ctggcaactactcttgtaattgtcatttctaaagcggagcacacctgaacccgcaat 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2650 CTTGAGACACGCGCTTTTATTTGGTATCAAAAGAGTGGTACGCGCTTGAACACTCAGT 2591

QY 123 ggtctactagaagtacagaaagccttcagatcggggctgcaattctcgaaacacgggt 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2590 AATGACCTTTCCGTCGCACACGTCCTTCGTCACAGCAGTCTAGAACCCGACACACAGTTC 2531

QY 183 gttgcaattactcaagaataattctctgttgataacactgc 223
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2530 GTGGCAGTTACCGAGCAGACCGACTCCACTTGCATAATAATCC 2490

RESULT 8
AL365270/c      156995 bp      DNA      HTG      07-JUL-2000
LOCUS          Homo sapiens chromosome 1 clone RP11-99A8, *** SEQUENCING IN
DEFINITION     PROGRESS ***, 16 unordered pieces.
ACCESSION      AL365270
VERSION        AL365270.1 GI:8980271
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 156995)
AUTHORS        Pavlitt,R.
TITLE          Direct Submission
JOURNAL        Submitted (06-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
               requests: clonerequest@sanger.ac.uk
               ----- Genome Center
               Center: Sanger Centre
               Center code: SC
               Web site: http://www.sanger.ac.uk
               Contact: humquery@sanger.ac.uk
               ----- Project Information
               Center project name: BA99A8
               ----- Summary Statistics
               Assembly program: XGAP4; version 4.5
               Sequencing vector: plasmid; L08752; 100% of reads
               Chemistry: Dye-terminator ET-amersham; 18% of reads Chemistry:
               Dye-terminator Big Dye; 81% of reads
               Consensus quality: 148967 bases at least Q40
               Consensus quality: 152162 bases at least Q30
               Consensus quality: 154190 bases at least Q20
               Insert size: 155495; sum-of-contigs
               Insert size: 143559; agarose-fp
               Quality coverage: 4.30x in Q20 bases; sum-of-contigs Quality
               coverage: 4.65x in Q20 bases; agarose-fp
               -----
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 16 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *
               * 1 10300: contig of 10300 bp in length
               * 10301 10400: gap of 100 bp
               * 10401 13921: contig of 3521 bp in length
               * 13922 14021: gap of 100 bp
               * 14022 27962: contig of 13961 bp in length
               * 27963 28082: gap of 100 bp
               * 28083 30195: contig of 2117 bp in length
               * 30200 30299: gap of 100 bp
               * 30300 36658: contig of 6359 bp in length
               * 36659 36758: gap of 100 bp

```

```

* 36759 61477: contig of 24719 bp in length
* 61478 61577: gap of 100 bp
* 61578 75491: contig of 13914 bp in length
* 75492 75591: gap of 100 bp
* 75592 90911: contig of 15320 bp in length
* 90912 91011: gap of 100 bp
* 91012 99070: contig of 8059 bp in length
* 99071 99170: gap of 100 bp
* 99171 101293: contig of 2123 bp in length
* 101294 101393: gap of 100 bp
* 101394 109454: contig of 8061 bp in length
* 109455 109554: gap of 100 bp
* 109555 110702: contig of 1148 bp in length
* 110703 110802: gap of 100 bp
* 110803 112576: contig of 1774 bp in length
* 112577 112676: gap of 100 bp
* 112677 139381: contig of 26705 bp in length
* 139382 139481: gap of 100 bp
* 139482 148660: contig of 9179 bp in length
* 148661 148760: gap of 100 bp
* 148761 156995: contig of 8235 bp in length.
*
FEATURES             Location/Qualifiers
     source            1..156995
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone_lib="RPC1-11.1"
     misc_feature      1..10300
                     /note="assembly_fragment:00909"
                     fragment_chain:1"
                     10401..13921
                     /note="assembly_fragment:00428"
                     fragment_chain:1"
     misc_feature      14022..27982
                     /note="assembly_fragment:00957"
                     fragment_chain:1"
     misc_feature      28083..30199
                     /note="assembly_fragment:00712"
                     fragment_chain:2"
     misc_feature      30300..36658
                     /note="assembly_fragment:00682"
                     fragment_chain:2"
     misc_feature      36759..61477
                     /note="assembly_fragment:00988"
                     fragment_chain:3"
     misc_feature      61578..75491
                     /note="assembly_fragment:00212"
                     fragment_chain:3"
     misc_feature      75592..90911
                     /note="assembly_fragment:01372"
                     fragment_chain:4"
     misc_feature      91012..99070
                     /note="assembly_fragment:00923"
                     fragment_chain:4"
     misc_feature      99171..101293
                     /note="assembly_fragment:00031"
     misc_feature      101394..109454
                     /note="assembly_fragment:00341"
     misc_feature      109555..110702
                     /note="assembly_fragment:00691"
     misc_feature      110803..112576
                     /note="assembly_fragment:00755"
     misc_feature      112677..139381
                     /note="assembly_fragment:01073"
     misc_feature      139482..148660
                     /note="assembly_fragment:01141"
     misc_feature      148761..156995
                     /note="assembly_fragment:00305"
                     clone_end:SP6
                     vector_side:right"
BASE COUNT          45192 a 31015 c 30671 g 48611 t 1506 others
ORIGIN

```

```

Query Match      11.9%; Score 35.6; DB 85; Length 156995;
Best Local Similarity 52.7%; Pred. NO. 4.5;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 120 aatggtctactagaatcacagaacccctcagagatcggggctgcaattctcgaacag 179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46917 AAAGGTTTCAAGGAGATGGAACGATATAGAGAAATCTGAAGGTCACTAAACAAG 46858
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 ggtgttgcaattactcaagaataattctctgttgataacactgctagaatagaggatg 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46857 GAAAGACCAATAGAAATTGTAATATCCAGGTAAATAAATAGAGTATTATTGCCCTCT 46798
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 240 ttatctcggtttctctatgttgattg 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46797 TAAGTTGTTATAATATGATGATTG 46772
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AL353604/c
LOCUS      AL353604      101370 bp      DNA      HTG      09-AUG-2000
DEFINITION Homo sapiens chromosome 1 clone RP5-1180C18 map p32.3-34.2, ***
SEQUENCING IN PROGRESS ***, 2 unordered pieces.
ACCESSION      AL353604
VERSION      AL353604.9 GI:9795005
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Wall, M.
Direct Submission
Submitted (08-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 11, 2000 this sequence version replaced gi:9717057.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj1180C18
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ABI; 3% of reads
Chemistry: Dye-terminator Big Dye; 45% of reads
Chemistry: Dye-terminator ET-amersham; 50% of reads
Quality: 99056 bases at least Q40
Consensus quality: 100022 bases at least Q30
Consensus quality: 100489 bases at least Q20
Insert size: 101270; sum-of-contigs
Insert size: 114210; 4.1% error; agarose-fp
Quality coverage: 0.00x in Q20 bases; sum-of-contigs Quality
coverage: 0.00x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 76800: contig of 76800 bp in length
*
* 76801 76900: gap of 100 bp
*
* 76901 101370: contig of 24470 bp in length.
      Location/Qualifiers
        1..101370
          /organism="Homo sapiens"

/db_xref="taxon:9606"
/chromosome="1"
/map="p32.3-34.2"
/clone="RP5-1180C18"
/clone_lib="RPCI-5"
1..76800
/note="assembly_fragment:00905.0"
76901..101370
/note="assembly_fragment:01232.0"

BASE COUNT 24879 a 24116 c 25164 g 27111 t 100 others
ORIGIN

Query Match      11.8%; Score 35.4; DB 64; Length 101370;
Best Local Similarity 49.7%; Pred. NO. 4.9;
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 120 aatggtctactagaatcacagaacccctcagagatcggggctgcaattctcgaacag 179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60814 AATGCGATGAAGGAGTCAAGGAAGATGCCACAGATGCTGGCCTGAGAACTGAAAAAT 60755
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 180 ggtgttgcaattactcaagaataattctctgttgataacactgctagaatagaggatg 239
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60754 GGAGCTTCTGTAACCAATAAATTTACCTTTTATGTTTAAAGCCAGTTGAACTAGGGTTA 60695
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 240 ttatctcggtttctctatgttgattggttggtggtgacttcagaaatgtctgctgtg 299
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60694 TCACTTGGAAATTTTTTTTAAATATTGAGTAATTCAGAGACTAACCAACAGCCAGAGGTA 60635
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 300 g 300
Db 60634 G 60634

RESULT 10
AL353604/c
LOCUS      AL353604      11836 bp      DNA      BCT      26-MAY-2000
DEFINITION Chlamydia muridarum, section 83 of 85 of the complete genome.
ACCESSION      AE002356 AE002160
VERSION      AE002356.2 GI:8163342
KEYWORDS
SOURCE      Chlamydia muridarum.
ORGANISM      Chlamydia muridarum
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE      1 (bases 1 to 11836)
AUTHORS      Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F.,
              White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
              Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,
              Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,
              McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
              Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
              pneumoniae AR39
              Nucleic Acids Res. 28 (6), 1397-1406 (2000)
REFERENCE      2 (bases 1 to 11836)
AUTHORS      Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F.,
              White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
              Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,
              Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J.,
              McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
              Direct Submission
              Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
              Medical Center Dr, Rockville, MD 20850, USA
              On Jun 1, 2000 this sequence version replaced gi:7190921.
              Location/Qualifiers
                1..11836
                  /organism="Chlamydia muridarum"
                  /db_xref="taxon:83560"
                  /note="synonym: Chlamydia trachomatis MoPn"
                  complement(69..1301)
                  /gene="TC0894"
                  complement(69..1301)
                  /gene="TC0894"

FEATURES
source
gene
CDS

```


/note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF39688.1"
 /db_xref="GI:7190922"
 /translation="MKVYCKLVVMVALLFSPSMGHALVQVGLERIFQBEKYLEKTRGRRI
 ALISHAAINQGEHSICVFDQHGKICLSALCTLEHGFAGASIAETPGYDPVLAGIH
 VLSFSKEIPSEVIEACDVFVDQDVGVSYSFISALIQVVKASASNKELIVLDR
 PNMGGDWGDLPEKALPAICYGWTGPELALLYRAWAPNATVTVVPMOGWTRS
 MTFADGRVWPSQVPPDQASAFYATGVLSTNIGICYTLPEKVLGAPWIDG
 HKVAQELNKAFLPKVKEPLPEMPEFFGKFMKEICSGVLLVLODPKTELPMTQSVILG
 VKMLYPKEVEQAFLLDLRLAPRRKAIQTLLGRSEFLNVCLOKQIVTWPLRTMCVEGR
 KQFKEARPFLLAETQQ"
 1496..2125
 /gene="TC0895"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF39689.1"
 /db_xref="GI:7190923"
 /translation="MKLIIASSHGYKVRKTKAFUKKIGEDIFSLVDYPTPKETG
 ETPEENATQKGVFAAQTFRCTWIADDSMLIIPALGGLPGKLSAFSGEHASDKDKRRK
 LLEMLLENIDRSAYPECCVWLVPFKIFKAHSCGTIVFKERSSGFGYDPLF
 SKHDYKQIVAELEPKEIKNOVSHRAKALAKLPYVEMAFANHILLARNESL"
 complement(2099..2326)
 /gene="TC0896"
 complement(2099..2326)
 /gene="TC0896"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF39690.1"
 /db_xref="GI:7190924"
 /translation="MIDGQTCFSGATHRLTAKSTVSLMPLATHNLQEGASAKLE
 ADFIRAEQTLAEQMQRSLQSLQSLTLIPRE"
 complement(2323..3012)
 /gene="TC0897"
 complement(2323..3012)
 /gene="TC0897"
 /note="similar to GB:M20681 SP:P11169 PID:306821;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="uracil-DNA glycosylase"
 /protein_id="AAF39691.1"
 /db_xref="GI:7190925"
 /translation="MHEAFTIEQLPSWQEQKLDWSPQYWSQLTLFLKSEYANATVY
 PKHEVFSALRSTPFDQVNVILGQDPYHGEQAGLSFVPKGOALPPSLRNIPQEL
 QADLGRNETGCLQWADQGVLLNTVLTVRAGEAFSHAGRGWEHFTDAIVTKLQNR
 THIFVWGSAAKKCDLLEFQTKHQHAILACHPSPFLAHRGFFCCCHFSKINYLKK
 QCKPMINWKIA"
 3093..4997
 /gene="TC0898"
 3093..4997
 /gene="TC0898"
 /note="similar to GB:M63176 PID:153062 SP:Q53727;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="ATP-dependent helicase PcrA"
 /protein_id="AAF39692.1"
 /db_xref="GI:7190926"
 /translation="MLTSELNRAQVTAATAPLOVLVLGAGAGKTRVYSHRILYLIE
 EAQLDPSQILAITFTTKAAKELNERVLTQCNFPDYGIPWSTFSLGLVYILRRSQ
 LDRQSNFAIYDQSDSEKLIKQCLRLNKLNLKCNMGQFTISQAKNKLQPEDLDSREY

PPARVYTEYOELRAANALDFFDLFLTEKLLRIPEIQOEYSNLKALLIDYQDT
 NHAQVLIARLAASHNNFVVGDPDOSTYSWRGANISNLFQDYFPAKALVRLSENY
 RSCGTILEANALTONNSARLDKTLRSVKGPDKLFCTGKNDDEAEQVLEESNLH
 SYKDIPLSHCILCYTQSQSFEAALLKRGYPYEIIGGISFYKRKEIQDILALFLRF
 SNYDMAAFERTIGLKCGIGATTLSLSYAKSMWDPILQACMDVLEKFRIRLTQKQ
 QOGLFSYLTHFQMEKLYNGCDLHEFINETITDYLAILEDKEDPETYDRKNLEQLL
 ENTLHWGSSNLPNLFEDLAKSADETASSHDRLKMTIHNSKGLFFPVVVLVLE
 AATLWKSANSGMHNIEERLCYVGITRAQEYLYLSRAKTRFLWGTERTVWPSRFTS
 ELPRALLAFV"
 5001..6281
 /gene="TC0899"
 5001..6281
 /gene="TC0899"
 /note="RNA polymerase sigma factor, sigma-54 family;
 identified by match to PFAM protein family HMM PF00309"
 /codon_start=1
 /transl_table=11
 /product="RNA polymerase sigma factor, sigma-54 family"
 /protein_id="AAF73614.1"
 /db_xref="GI:8163343"
 /translation="MLHQTASVALCPALHIOQSLDMLQMPVAELSEFLSQITINP
 CFDLALDLPDTCSEFFPISSEHHPETETLISLLHQIEVHPSLPDRAIAOVIVGNLS
 PEGFLDEPSLSASNLGVSDVLFQKILWRIQOFHPLGATSSLSQSWLSLLOPSSHKE
 ALMIIRQFNRLARCDFAGISKKIQIOLPAKILVFLRQALASTPNCGPAFPNSLTKPS
 PALPDAYLSFSHEKSWETPMNKDSLPAIRLNSVLIHYHTLPRDDKHLSSQIRAAKQ
 LLNRINKRETELAVLRVLIPYQEEFLKRSVPKAFSVKQLARELSLHETTCRAIN
 NKTATPGLISMSRLFPQAVGACPDQSKATILHWHQIWISTEKNPLSDEAISQIIA
 KGIPCARRVAKYRSQNLNPPAHQRKHNASLVP"
 complement(6405..7115)
 /gene="TC0900"
 complement(6405..7115)
 /gene="TC0900"
 /note="similar to SP:P27505 PID:43907; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="coenzyme PQQ synthesis protein c, putative"
 /protein_id="AAF39693.1"
 /db_xref="GI:7190927"
 /translation="MESRKGIKSVSMNFDQLDQDAITQNKHMLEHPHYMKWSKGLTKE
 QLOAYAKDYVLIHKAFKYLKLSARCDDELEAKKLLDLNDEENGYPNHDLMKQFV
 FALGVSEELAEHPSEAAKAVATMRCTGDSLAAGVAALYSVESQIPCVAKEKIR
 GLIEYFGSPEDYATITHEEADVRHAREEKALIELMSRDDSDKVKLEASREVTSQSLY
 GFLDSFLEPATCCHKA"
 complement(7094..7825)
 /gene="TC0901"
 complement(7094..7825)
 /gene="TC0901"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF39694.1"
 /db_xref="GI:7190928"
 /translation="MKITPKTKRVFAHDSLQELFOESLPPIQERSITVVVSSKIVSIC
 KGVADAOIKSADLILKEADAYLFCESGILYTKKGGILIPSIDESNTDQAFVLP
 EDALRSCNQIGEWLKNVFRYKELGVITDSTTPMRGVVVGICLWGFSPHNYVGS
 LDCDFRLKMTQSNLLDLSAAAVCMGEGNEQTPLAVIEQAPKVVVYHOHPTSOEYF
 SLRIDETEDLYGPLLHVLAVLNOEKE"
 complement(7797..8276)
 /gene="TC0902"
 complement(7797..8276)
 /gene="TC0902"
 /note="similar to GB:D10483 SP:P00379 GB:J01609 GB:V00276
 GB:X05108; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="dihydrofolate reductase"
 /protein_id="AAF39695.1"
 /db_xref="GI:7190929"
 /translation="MIQATGIVADIPRGVMGAGKLPWNPEDIRFFSETIODHPILM
 GRKTWESLPDRYKCGRTVIVFSRQHSQAQGIWISLVEYEKLSLNSPFLIGGAEFLW


```

FFOYNLLKSCFVTHIKREVQDTPFPVRLSGWKRESVLKTDENFIHYENYANQNP"
complement(8273.. .9625)
/genes="TC0903"
CDS
complement(8273.. .9625)
/genes="TC0903"
note="similar to PID:169414; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="2-amino-4-hydroxy-6-hydroxymethylidihydropterin
pyrophosphokinase/dihydropteroate synthase"
/protein_id="AAF39696.1"
/db_xref="GI:7190930"
translation="MTWTFVCLGSLNRRHEYIKRAYSLLKAGIRNLKSSVILE
TKALLGSPKEMDLPYCNVAGETOLSPDELKIKMIENLRSLKRWGPRSID
IDVLLGDESYSCSERCIIIPHLRLLPMLMASLCPYRFLRGLSPYDGKTFAE
LAAIYPTEKDLVGLSFAPPTQIMGVNVTDSITDGLFLKRAAAHAERLFAEGAS
Query Match 11.7% Score 35; DB 1; Length 11836;
Best Local Similarity 56.5%; Pred. No. 5.1;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 123 ggtctactagaagtacagaagccttcagagatcgccgggtcgcaattctcgaaacaggggt 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7121 GGTGTGTTGGTAAGAAAGGAATAAAGAGGTGACGATGATTTTATAGATCAGCTA 7062
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 183 gttgcaattactcaagaatactctgttgataacactctagaatagaggat 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7061 GATGCAATTATTCRAAACACATATGTTAGACACCCCTTTTACATGAAGTGTG 7007
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AC023104/c
LOCUS
DEFINITION
AC023104 Homo sapiens chromosome X clone RP11-733018 map X, WORKING DRAFT
VERSION
AC023104.2 GI:7139841
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162149)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguski,M., Buckler,A., Brown,A., Burkett,G., Castile,A.,
Choepe,Y., Collinge,M., Collins,S., Collins,S., Collymore,A., Cooke,P.,
DeArrellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2000 this sequence version replaced gi:6939348.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
FEATURES
source
misc_feature
1..1573
Location/Qualifiers
1..162149
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone_lib="RP11-733018"
/clone_lib="RP11-11 Human Male BAC"
1..1573

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6728
Center clone name: 733 O.18
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 150591 bases at least Q40
Consensus quality: 155557 bases at least Q30
Consensus quality: 158001 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 160349; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1573: contig of 1573 bp in length
1574 1673: gap of 100 bp
1674 3543: contig of 1870 bp in length
3544 3643: gap of 100 bp
3644 5601: contig of 1958 bp in length
5602 5701: gap of 100 bp
5702 7732: contig of 2031 bp in length
7733 7832: gap of 100 bp
7833 10667: contig of 2835 bp in length
10668 10767: gap of 100 bp
10768 13917: contig of 3150 bp in length
13918 14017: gap of 100 bp
14018 17381: contig of 3364 bp in length
17382 17481: gap of 100 bp
17482 20415: contig of 2934 bp in length
20416 20515: gap of 100 bp
20516 23137: contig of 2622 bp in length
23138 23237: gap of 100 bp
23238 27089: contig of 3852 bp in length
27090 27189: gap of 100 bp
27190 33459: contig of 6270 bp in length
33460 33559: gap of 100 bp
33560 39422: contig of 5863 bp in length
39423 39522: gap of 100 bp
39523 46446: contig of 6924 bp in length
46447 46546: gap of 100 bp
46547 56042: contig of 9496 bp in length
56043 56142: gap of 100 bp
56143 67767: contig of 11625 bp in length
67768 67867: gap of 100 bp
67868 86875: contig of 19008 bp in length
86876 86975: gap of 100 bp
86976 106760: contig of 19785 bp in length
106761 106860: gap of 100 bp
106861 132195: contig of 23335 bp in length
132196 132295: gap of 100 bp
132296 162149: contig of 29854 bp in length.

FEATURES

source

misc_feature

Sat Nov 4 18:11:41 2000

```
misc_feature /note="assembly_fragment"  
1674. .3543  
misc_feature /note="assembly_fragment"  
3644. .5601  
misc_feature /note="assembly_fragment"  
5702. .7732  
misc_feature /note="assembly_fragment"  
7833. .10667  
misc_feature /note="assembly_fragment"  
10768. .13917  
misc_feature /note="assembly_fragment"  
14018. .17381  
misc_feature /note="assembly_fragment"  
clone_end:T7  
vector_side:right  
17482. .20415  
misc_feature /note="assembly_fragment"  
20516. .23137  
misc_feature /note="assembly_fragment"  
23238. .27089  
misc_feature /note="assembly_fragment"  
clone_end:SP6  
vector_side:left  
27190. .33459  
misc_feature /note="assembly_fragment"  
33560. .39422  
misc_feature /note="assembly_fragment"  
39523. .46446  
misc_feature /note="assembly_fragment"  
46547. .56042  
misc_feature /note="assembly_fragment"  
56143. .67767  
misc_feature /note="assembly_fragment"  
67868. .86875  
misc_feature /note="assembly_fragment"  
86976. .106760  
misc_feature /note="assembly_fragment"  
106861. .132195  
misc_feature /note="assembly_fragment"  
132296. .162149  
misc_feature /note="assembly_fragment"  
50445 a 30230 c 30351 g 49322 t 1801 others  
BASE COUNT  
ORIGIN  
Query Match 11.7%; Score 35; DB 48; Length 162149;  
Best Local Similarity 61.5%; Pred. No. 6.8;  
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
QY 207 ctgttgataacactgtaagaaatagagggatggttagctcggttctctatgtttgattgg 266  
Db 116844 CTGTTAAATTAATAATGCTACACATCATGTTCTGTTTACCTGTTTATTTATTTTCATTGG 116785  
QY 267 gttgtgttaggacttcagaaatgtctgtc 297  
Db 116784 ATTTTGGTATGACTATACATGATCTGCTG 116754  
RESULT 12  
AC019951/c  
LOCUS  
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered  
pieces  
ACCESSION AC019951  
VERSION AC019951.1 GI:6664946  
KEYWORDS HTG; HTGS_PHASE2  
SOURCE fruit fly  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 33328)  
AUTHORS Adams, M. and Venter, J. C.
```

```
Direct Submission  
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT  
This sequence was identified as CDM:10211508 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
* NOTE: This is a 'working draft' sequence.  
* This sequence will be replaced  
* by the finished sequence as soon as it is available and  
* the accession number will be preserved.  
FEATURES  
Source  
1. .33328  
/organism="Drosophila melanogaster"  
/db_xref="taxon:7227"  
BASE COUNT 9249 a 7305 c 7379 g 9395 t  
ORIGIN  
Query Match 11.5%; Score 34.6; DB 40; Length 33328;  
Best Local Similarity 49.2%; Pred. No. 7.6;  
Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
QY 94 agagcggaggcacacacacccgcaatggtctctacaaagtacagaaagccttcagag 153  
Db 12366 ATAGCATTTGAATCATTCACACCTGCTATTTTATAGCGGAATACAAAGCTCTTCTGTG 12307  
QY 154 atcggggctgcaattctcgaaacagggtgtgcaattactcaagaaatctctctgtg 213  
Db 12306 ATACGGATATATGCTATACATTTGAGGCTTTTATATCTACTAGATAAATTTTAAATAA 12247  
QY 214 ataacactgctagaatagaggatggttagctcggttctctatgtttgattgggttg 273  
Db 12246 GTCTGCTTACAAATAGATCCATAATTCATGCTGGTGTGTTTTTAACTCTGTAGTCGCG 12187  
QY 274 gtagg 278  
Db 12186 CGAGG 12182  
RESULT 13  
AC010575  
LOCUS  
DEFINITION Drosophila melanogaster clone RPC198-5W7, *** SEQUENCING IN  
PROGRESS ***, 52 unordered pieces.  
ACCESSION AC010575  
VERSION AC010575.3 GI:6996792  
KEYWORDS HTG; HTGS_PHASE1  
SOURCE fruit fly  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE  
1 (bases 1 to 148447)  
AUTHORS Muzny, D. M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,  
Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
Burkett, C., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K. J., Fernandez, C., Ferraguto, D.,  
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J. H., Hoques, M.,  
Guevara, W., Harris, K., Hernandez, J., Hodgson, L., Jia, Y., Jones, M.,  
Holloway, C., Hosak, H., Jackson, L. E., Jackson, L., Leal, B., Li, Z.,  
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
Lichter, O., Liu, J., Liu, W., Logan, O., Lozano, R. J., Lu, J.,  
Lucier, R., Martin, R., Martinez, C., McLeod, M. P., Mei, G., Morgan, M.,  
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,  
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L. L.,  
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,  
Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugranger, R.,  
Tabor, P., Taylor, T., Vasquez, I., Vinson, R., Vo, Q., Wahbah, M.,  
Watlington, S., Weinstein, G., Weinstein, I. R., Williamson, A.,  
Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and  
Gibbs, R.  
Direct Submission  
Unpublished  
TITLE  
JOURNAL
```

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 148447)

Worley, K.C.

Direct Submission

Submitted (16-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 18, 2000 this sequence version replaced gi:5902992.

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: DRGW

Center clone name: RCI98-5M7

----- Summary Statistics

Sequencing vector: M13; L08821

Chemistry: Dye-terminator Big Dye; 17% of reads

Assembly program: Phrap; version 0.980611

Consensus quality: 105301 bases at least Q40

Consensus quality: 119508 bases at least Q30

Consensus quality: 126701 bases at least Q20

Estimated insert size: 132501; sum-of-coverage estimation

Quality coverage: 0.6x in Q20 bases; sum-of-coverage estimation

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 52 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 850: contig of 850 bp in length
 * 851: gap of unknown length
 * 871: 1717: contig of 847 bp in length
 * 1718: 1737: gap of unknown length
 * 1738: 2614: contig of 877 bp in length
 * 2615: 2634: gap of unknown length
 * 2635: 3454: contig of 820 bp in length
 * 3455: 3474: gap of unknown length
 * 3475: 4393: contig of 919 bp in length
 * 4394: 4413: gap of unknown length
 * 4414: 5289: contig of 876 bp in length
 * 5290: 5309: gap of unknown length
 * 5310: 6141: contig of 832 bp in length
 * 6142: 6161: gap of unknown length
 * 6162: 6984: contig of 823 bp in length
 * 6985: 7004: gap of unknown length
 * 7005: 8080: contig of 1076 bp in length
 * 8081: 8100: gap of unknown length
 * 8101: 9327: contig of 1227 bp in length
 * 9328: 9347: gap of unknown length
 * 9348: 10228: contig of 881 bp in length
 * 10229: 10248: gap of unknown length
 * 10249: 11302: contig of 1054 bp in length
 * 11303: 11322: gap of unknown length
 * 11323: 12174: contig of 852 bp in length
 * 12175: 12194: gap of unknown length
 * 12195: 13297: contig of 1103 bp in length
 * 13298: 13317: gap of unknown length
 * 13318: 14539: contig of 1222 bp in length
 * 14540: 14559: gap of unknown length
 * 14560: 15698: contig of 1139 bp in length
 * 15699: 15718: gap of unknown length
 * 15719: 16522: contig of 804 bp in length
 * 16523: 16542: gap of unknown length
 * 16543: 17518: contig of 976 bp in length
 * 17519: 17538: gap of unknown length
 * 17539: 18404: contig of 866 bp in length
 * 18405: 18424: gap of unknown length
 * 18425: 19389: contig of 965 bp in length
 * 19390: 19409: gap of unknown length

19410 20870: contig of 1461 bp in length
 * 20871 20890: gap of unknown length
 * 20891 22733: contig of 1845 bp in length
 * 22736 22755: gap of unknown length
 * 22756 24425: contig of 1670 bp in length
 * 24426 24445: gap of unknown length
 * 24446 25358: contig of 913 bp in length
 * 25359 25378: gap of unknown length
 * 25379 26724: contig of 1346 bp in length
 * 26725 26744: gap of unknown length
 * 26745 28410: contig of 1666 bp in length
 * 28411 28430: gap of unknown length
 * 28431 29885: contig of 1455 bp in length
 * 29886 29905: gap of unknown length
 * 29906 30781: contig of 876 bp in length
 * 30782 30801: gap of unknown length
 * 30802 31715: contig of 914 bp in length
 * 31716 31735: gap of unknown length
 * 31736 32774: contig of 1039 bp in length
 * 32775 32794: gap of unknown length
 * 32795 34403: contig of 1609 bp in length
 * 34404 34423: gap of unknown length
 * 34424 35892: contig of 1469 bp in length
 * 35893 35912: gap of unknown length
 * 35913 37899: contig of 1987 bp in length
 * 37900 37919: gap of unknown length
 * 37920 39450: contig of 1531 bp in length
 * 39451 39470: gap of unknown length
 * 39471 41240: contig of 1770 bp in length
 * 41241 41260: gap of unknown length
 * 41261 43877: contig of 2617 bp in length
 * 43878 43897: gap of unknown length
 * 43898 45845: contig of 1948 bp in length
 * 45846 45865: gap of unknown length
 * 45866 48533: contig of 2668 bp in length
 * 48534 48553: gap of unknown length
 * 48554 50809: contig of 2256 bp in length
 * 50810 50830: gap of unknown length
 * 50831 55073: contig of 4224 bp in length
 * 55074 55073: gap of unknown length
 * 55074 58837: contig of 3764 bp in length
 * 58838 58857: gap of unknown length
 * 58858 62191: contig of 3334 bp in length
 * 62192 62211: gap of unknown length
 * 62212 65883: contig of 3672 bp in length
 * 65884 65903: gap of unknown length
 * 65904 70458: contig of 4555 bp in length
 * 70459 70478: gap of unknown length
 * 70479 75989: contig of 5511 bp in length
 * 75990 76010: contig of 6481 bp in length
 * 82491 82510: gap of unknown length
 * 82511 88156: contig of 5646 bp in length
 * 88157 88176: gap of unknown length
 * 88177 94440: contig of 6264 bp in length
 * 94441 94460: gap of unknown length
 * 94461 100652: contig of 6192 bp in length
 * 100653 100672: gap of unknown length
 * 100673 110078: contig of 9406 bp in length
 * 110079 110098: gap of unknown length
 * 110099 121746: contig of 11648 bp in length
 * 121747 121766: gap of unknown length
 * 121767 148447: contig of 26681 bp in length.

FEATURES
source

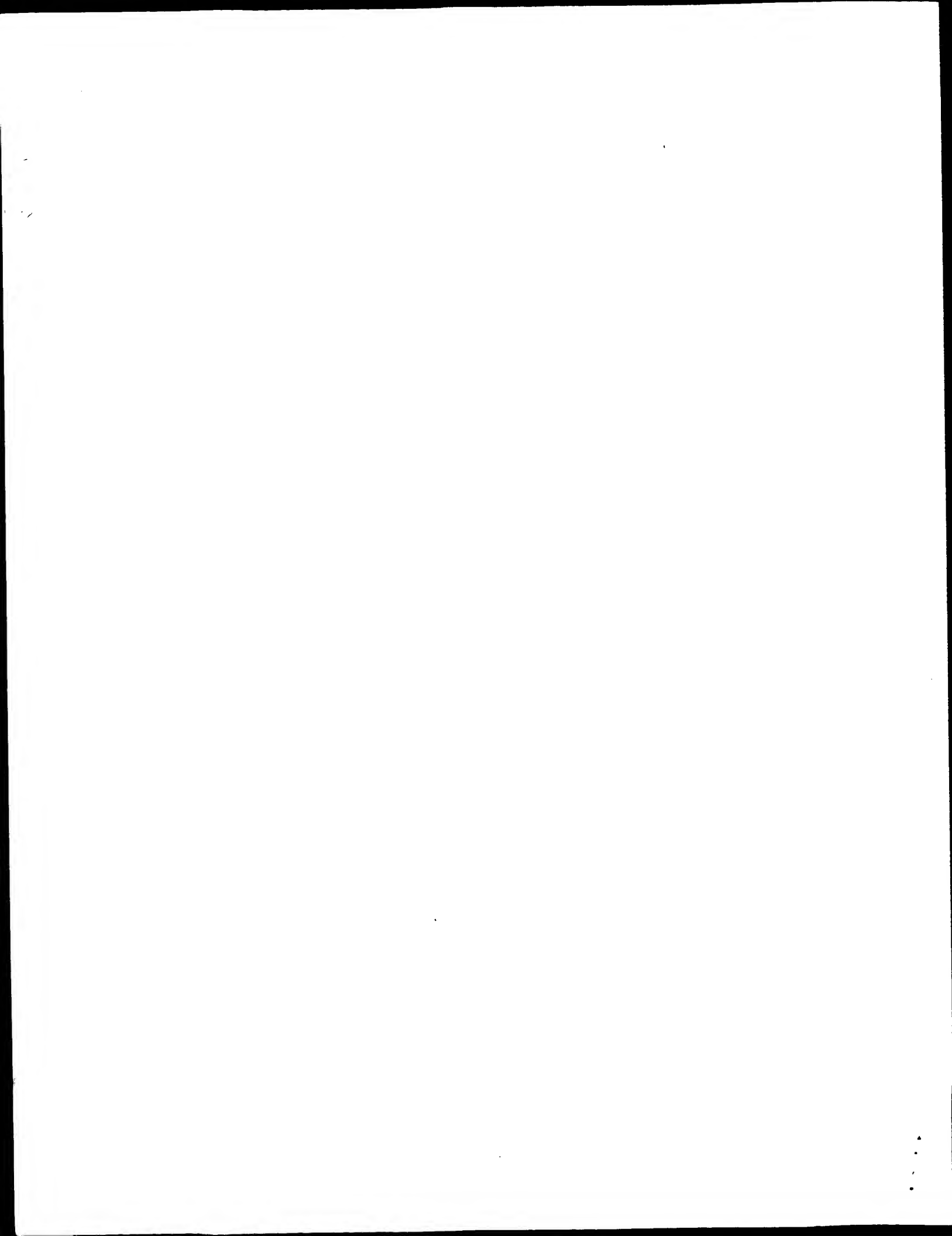
1. 148447
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="RCI98-5M7"

BASE COUNT
ORIGIN

42418 a 32481 c 31053 g 41428 t 1067 others

Query Match

11.5%; Score 34.6; DB 36; Length 148447;



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:32:00 ; Search time 4075.18 Seconds
(without alignments)
245.431 Million cell updates/sec

Title: US-09-300-482-27

Perfect score: 229

Sequence: 1 cagacctattttttctgc.....cttcaatttatggggaacaa 229

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba1:

2: gb_ba2:

3: gb_om:

4: gb_ov:

5: gb_pat:

6: gb_ph:

7: gb_pll:

8: gb_pl2:

9: gb_pr1:

10: gb_pr2:

11: gb_pr3:

12: em_fun:

13: em_hum1:

14: em_hum2:

15: em_in:

16: em_om:

17: em_or:

18: em_ov:

19: em_pat:

20: em_ph:

21: em_pl:

22: em_ro:

23: em_sts:

24: em_sy:

25: em_un:

26: em_vi:

27: gb_htg1:

28: gb_htg2:

29: gb_in1:

30: gb_in2:

31: em_ba1:

32: em_ba2:

33: em_hum3:

34: em_hum4:

35: gb_pr4:

36: gb_htg3:

37: gb_htg4:

38: gb_htg5:

39: gb_htg6:

40: gb_htg7:

41: em_htg1:

42: em_htg2:

43: em_htg3:

44: em_hum5:
45: gb_pl3:
46: gb_pr5:
47: gb_htg8:
48: gb_htg9:
49: gb_htg10:
50: gb_htg11:
51: gb_htg12:
52: gb_htg13:
53: gb_htg14:
54: gb_in3:
55: gb_htg15:
56: gb_htg16:
57: gb_htg17:
58: em_htg4:
59: em_htg5:
60: em_htg6:
61: em_htg7:
62: em_hum6:
63: gb_htg18:
64: gb_htg19:
65: gb_ba3:
66: em_htg8:
67: em_htg9:
68: em_htg10:
69: em_htg11:
70: em_htg12:
71: em_htg13:
72: em_htg14:
73: em_htg15:
74: em_htg16:
75: em_htg17:
76: em_htg18:
77: em_htg19:
78: em_htg20:
79: em_htg21:
80: em_htg22:
81: em_htg23:
82: gb_pr6:
83: gb_pr7:
84: gb_htg20:
85: gb_htg21:
86: gb_htg22:
87: gb_htg23:
88: gb_ro:
89: gb_sts1:
90: gb_sts2:
91: gb_sy:
92: gb_un:
93: gb_vil:
94: gb_vil2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	136.2	59.5	1667	7	AB007907 Glycine m
2	84.2	36.8	1618	45	U18239 Medicago sa
3	77	33.6	96240	56	AC068900 Arabidops
4	60.2	26.3	634	7	AF037030 Zea mays
5	60.2	26.3	150120	8	AF001552 Oryza sat
6	58.6	25.6	1686	7	AF061838 Zea mays
7	57.4	25.1	1818	7	AF061837 Zea mays
8	57.4	25.1	2392	7	AF037037 Zea mays
9	55.8	24.4	645	7	AF037029 Zea mays
10	52.4	22.9	111222	7	AC007764 Genomic s
11	50.8	22.2	79837	7	AB005233 Arabidops
12	38.2	16.7	170453	10	AC006600 Homo sapi


```

/db_xref="taxon:56147"
/clone="019"
/tissue_type="cell suspension culture"
7.11467
/codon_start=1
/product="6-phosphogluconate dehydrogenase"
/protein_id="AAB41553.1"
/db_xref="GI:603221"
/translation="MAPPLTRIGLAGLVNQNLAIADKGFPISVNRTTSKVDET
VERAKOENPLPGFHDPEAFVNSIEKPRVIMLVKAGAPVEQTTKLSAYLEKDCI
IDGNEWENIEREKEVAELGILMGVSGEGEARGPMPGSGSFAYKYEDI
LLKAAVDPGSCVTVYIGNGSGNFVKMHNGIEYGDMLIAEAYDVLKSVKLTNE
ELQSAFTWNKGLLSFLIEITADIFGIKDDKGDVLDVKDLDTGKGTCKWTFQQA
AELSVPATTEASLDARFLSGMLDERVEAKVFKSGRGLDITDOQVKKQLIDDVKK
ALVAAKICVAQGMNLIIRAKSAEKWDLALGELARIWKGCCIIIFLDRIKQAYDRN
PNLANLLVDFEFAKEIIEIQTARVRVSVNSGSLSPGMSLASLAFDYSRRERLPAN
LVOAQRDYGAHTYERVDIEGSHTEWFKLAKQSI"
1618
polyA_site
BASE COUNT 461 a 284 c 405 g 468 t
ORIGIN

Query Match
Best Local Similarity 36.8%; Score 84.2; DB 45; Length 1618;
Matches 121; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

Qy 53 gctcaaccacaagaatagccttgcctgctggttaataaggcgaataatctggca 112
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 GCTCACCACCTTACAGATAGGCCCTGCGGATTCGGCTTATGGCCAGAAATCTCGCA 69
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 113 ctcaattctgtgaaaggcttcccaatccggttaacacggaacatttccaaagt 172
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70 CTCAATATTGC-CGACAAAGGATTCCTCAATTTCTGTTTATATAGAACACATCAAGGT 128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 173 tattggccataagacgaacaaacagggaagaacatttcaatt 217
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129 TGACGAACCTGTGACGAGCAACACAGAGGAGGAAATCTTCCACT 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
AC068900 96240 bp DNA HTG 02-JUN-2000
LOCUS
DEFINITION
Arabidopsis thaliana chromosome III clone IGF-F11A12. ***
SEQUENCING IN PROGRESS ***, 17 unordered pieces.
ACCESSION
AC068900
VERSION
AC068900.2 GI:8192588
KEYWORDS
HTG; HTGS.PHASE1.
SOURCE
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 96240)
Lin.X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Ronning,C.M.,
Koo,H., Fujii,C.Y., Utterback,T., Barnstead,M.E., Bowman,C.L.,
White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana 'IGF' BAC 'F11A12' genomic sequence near
marker 'mi174'
Unpublished
2 (bases 1 to 96240)
Lin.X. and Kaul,S.
Direct Submission
Submitted (11-MAY-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
On Jun 2, 2000 this sequence version replaced gi:7769972.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will

```

```

* be preserved.
* 1 22087: contig of 22087 bp in length
* 22088 22137: gap of unknown length
* 22138 25099: contig of 2962 bp in length
* 25100 25149: gap of unknown length
* 25150 27221: contig of 2072 bp in length
* 27222 27271: gap of unknown length
* 27272 29774: contig of 2503 bp in length
* 29775 29824: gap of unknown length
* 29825 34383: contig of 4559 bp in length
* 34384 34433: gap of unknown length
* 34434 37434: contig of 3001 bp in length
* 37435 37484: gap of unknown length
* 37485 41731: contig of 4247 bp in length
* 41732 41781: gap of unknown length
* 41782 43856: contig of 2075 bp in length
* 43857 43906: gap of unknown length
* 43907 46860: contig of 2954 bp in length
* 46861 46910: gap of unknown length
* 46911 50824: contig of 3914 bp in length
* 50825 50874: gap of unknown length
* 50875 56713: contig of 5839 bp in length
* 56714 56763: gap of unknown length
* 56764 61236: contig of 4472 bp in length
* 61236 61286: gap of unknown length
* 61286 66859: contig of 5574 bp in length
* 66860 66909: gap of unknown length
* 66910 73521: contig of 6612 bp in length
* 73522 73572: gap of unknown length
* 73572 78794: contig of 5222 bp in length
* 78794 78843: gap of unknown length
* 78844 83051: contig of 4208 bp in length
* 83052 83101: gap of unknown length
* 83102 96240: contig of 13139 bp in length.

FEATURES
Location/Qualifiers
1..96240
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="III"
/clone="IGF-F11A12"

BASE COUNT 28496 a 18610 c 20275 g 28013 t 846 others
ORIGIN

Query Match
Best Local Similarity 33.6%; Score 77; DB 56; Length 96240;
Matches 136; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

Qy 7 ttatttttctgcttgccttcaaatctcaggagagattaattatgcgtcaaccacaac 66
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12986 TGATGTATTATTGTTGGCAGAGAGAGAGACTTAGAAAAATGGCTTCAACCTAC 13045
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 67 aagaataggccttgcctggttgctggttaattggcgaataatgcgtcaaccacaac 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13046 AAGAATAGTCTAGCTGACCTGCTGTGATGGTCAGAGACCTAGCTCTCAACATTGCA-G 13104
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 127 aaaaagggttcccaatccggttaacacggaacatttccaaaggttattgggccaag 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13105 AGAAAGGCTTCCCAATCTCTGTTTACACAGAACTACTTCCAAAGTTGATGAGACCGTCG 13164
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 187 acgaagcaaacagggaagaaaccttcaattatggg 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13165 AACGAGCCAAAAGGAGGCAACTCTCTCTTTTACGG 13201
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
AF037030
LOCUS
DEFINITION
Zea mays 6-phosphogluconate dehydrogenase isoenzyme B mRNA, partial
cds.
ACCESSION
AF037030
VERSION
AF037030.1 GI:3925224
KEYWORDS

```


DGDRRGGDSRRRRHIAATMRGAARARCEGEGVLTGVRPATRETTNGDAA
BLOTLREEMREVRGREGVGEREPPTAAMGLTGTRRTTAFDGGIKWRGRRAAA
SGRRSCGGGGRNDRGHGRERKEEGEGEAL"
Join(19096...19172,20123...20329,20389...20395)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA93017.1"
/db_xref="GI:7363273"
/translation="MOLEPTMHOCLVOWGDKIEIVPANRPOSCRRRLGTAAPTRRP
DILPSPESRRRTLPAASAAAFSPAPASGGRHSISGSGSALRLSGLRT"
21587...21769
/note="5' LTR"
Join(21843...22489,22572...23072,23168...24293)
/note="Similar to Arabidopsis thaliana DNA chromosome 4,
BAC clone F17A8; putative protein. (AL049482)"
/codon_start=1
/protein_id="BAA93018.1"
/db_xref="GI:7363274"
/translation="MEKSTALLFOKHEAKKNACDLPIPHVDEKSSIPAVNDAPIDED
IVVDEGEETEEDVDDPPDVVDEVSTETKEESLPYVDVDDLEHDPGLRVPIS
SFANDQDAARRGYILKPGCHLWAFNPPSRKIYKDRFRSVIWEHKYWIYESVDKDA
TFCFVCLFEGESGKEVTVGWHNNVWNAKALDKHYGGTSSDHNFAQENYINLVKKGCL
RYLLROGLAFRGHDETESNNGNFLELLKWLGNANANKVYVNNAPGNCILTSPIRI
QKQVIRCCSWETTSICELGEDEHYAILADESSDISHKEQLVCLRYDKLGGVCERF
LAVVAGTSLSQLTATIOSLLTSHLLTTOIHGGQYDGSNMKEGVKGCWLPNLFQ
QVRLNIVGVSCTRHMDLRVRAQKALKALNGEIASGSLNGOEMALARPGLDRWGS
HYKTIHLIDMTYDITREVLITIGKOPTOREDWPIIHAMVLAFESFEVFNALHMLYL
VTFNEFSNLQKRDODIVNAMSVLGLAKKQWOMEFNGEWELGKVSFCIKYSIDIP
AMDKVVPHGRSHRFYPTQIDHRYREYVIGVIERHQELENRFEDESMELLCSMA
FNPTDSFASDAQILRLASFYPKDIENSLMKLELDQDTYINDMREDRHFKGLNKIG
QISIKLVEFKKHDLDLYLLKLLILFLPVATASVERFVSAMNLVKTVRNMSDKLL
NCLVTFIERDMYMRKT"
26421...26603
/note="3' LTR"
complement(join(29217...30334,31321...31609,31701...31862))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA93019.1"
/db_xref="GI:7363275"
/translation="MYQIKIRKKRDEVLVATRLGKVESIQALRMDGRVSGDDAGE
RRPLRRACOGPKSKVLGDLWGLPWLITSLTHPPRLASTPIGRAAAAATAAGL
VSSPAHCKLRVSKRWNGAAGSGEDHGRRRRPSRSPSLAKGLMSMLHQI
IALLSSFDLLAFSGTCSWRAALSSPFIYTFPPLHLKPDIPNHPHCSFRYTL
LYKRWQGLDSKRLSKSAPONTNMYILGCSYGLIFSYENCLVDMYTAK
VYKPKLQAGNKEITYGILAPLPLSHLLCSRSSIFYWQVTSWSHPPGGRRI
LQVLFGEFTAMFPHRLHTMREAPQLSMQEVGVWGEEMFVHFKPLWISGDMIL
LMLDSVGIHHSYGFPGTFQVRLDFSAQTAKWMKLEKLSALFYSLDRNPFTSCT
NPERWKGKNCIYVAKPSEDSEDPWTAVELGQPIPGATHCVYPSHPLLRTEGHCQLE
YLWLPSPINGVDQ"
32671...33849
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA93020.1"
/db_xref="GI:7363276"
/translation="MQSPKVSSEASGAPPATSVIEGWAEPLPGLLHSIVALLGSFLD
LIAFTGCHSWRAAFSPYSPKSNFTLPLLYRPNVRVAKPISSHNGHRLKRSCEVID
LANRTPLCQIOPETLQRHFFAGSGGLICRRGTCVLDVDTGAEPVSPRLPSE
NCDREYCGILTAIPNSHLIIISTDPVSDSSELKLVNPNVDQVTFENG
QLIAVETSLTQALPILREKIKTLWMDNNECPMRPFVVCVGDMLLYDVHVISF
SFGAPVLYRLDMSTKPKARVKKLENWALFIGDARSPPSPFKPFRNGRSNC
LYAHYSQPLSLHGLGDDADAVDNPNTDNLVFRNRYQLQALWYPSMFTSADGGQ"
complement(join(34717...34935,35401...35912,36242...36363,
36470...36542,36626...36754,36936...37247,37336...38617,
38724...39212))
/note="Similar to Arabidopsis thaliana DNA chromosome V,
BAC clone TM021B04; N. tabacum membrane-associated
salt-inducible protein. (AF007271)"
/codon_start=1
/protein_id="BAA93021.1"
/db_xref="GI:7363277"
/translation="MAAATAGAAASTAVITGSSSYEDDGAATSWLSSTGSSSSSS
SQRPYRLLHDEAQLRRERRGSGAHTPRVRRTPDQMARVYEDDRAGHYVGRH

VVAARAARATASCSPPSSADMBREMASFVAKLTFRMCFVLHELRGWRQARDFPAM
KLQCYEPSVAVTILLRLQGVKVLAEVTFLEMLQAGCEPDVACGTGTLICAVARW
GKNDMLMFAAARRRDIPIVSTFNFVSSLOKOLHGVTHLMEQMLEANVAPNQF
TYTVVGSYAKGMELEAMDAFGEKRRFVPEATYSILLICAKHGKGEALGLYD
EMKVSIVSNYTCASVTLTYKNEDYSKALSLFSEMEONKIVPDEVYIGLVRIYK
LGLYEDAQRMFEEDIDKAGLSDQTYVMAQVMHNVQNDYALVLDAMRARNVPSQ
FSYALIRHCHVAKEDVDAEDTFRALSNGPDPCCNDLLRYLWRLGHLDKARALIL
MKREALQFQEDLCVTVLVCCKTSINKDTDNLTIVONEGSSSKVLPNTDSSTLSMM
KSLLDKPGGLSSVOLINKFREGSTDEAKELYEHLTEAGKPDPTATILIVYQGO
AQOLEAQKLFETASTSPVGGSVYNNAMDALCRGKTEEARLYFMEILDGNGDAV
TISILVHLTKQKFOEANIYVRCGLHDEALDVTYNTFIKMSLEGLDNVLLAVS
IDYRMLSSGIPRSMOTFNISYVGGQKLEKAVEMSAOELGIPIDETKVTNMLSF
YKAGLHNAEAIIFQEMKNHVPDSHTYLALIRAVTEGKCYSKAEATOMLRSNMT
PSTGHNLISAFLEKGOIDEAQMWNQMEAGIPADLACCTMRMHLDHGYVDDGI
LPFETARLLKPDSEFTLSAAFLHYEHSGRESAGDYLDANNNGASFURNLKVSKLE
QIOPKWSIVKSGPLPPOKAHGVPLKYKORSVCCSVEQVAREGEVDAARFGRSRA
AVHDHVQRVDSSTCD"
complement(join(40023...40413,40774...40971,41293...41464,
41783...42027,42111...42288,42719...42901,43197...43301,
44137...44274,44361...45353))
/note="ESTs C74776(E51022),C26123(C116681) correspond to a
region of the predicted gene.
Similar to Arabidopsis thaliana cultivar Landsberg
extra-large G-protein (AF060942)"
/codon_start=1
/protein_id="BAA93022.1"
/db_xref="GI:7363278"
/translation="NAGAGVGSNNWEEMVRMLPPGTTFEAPANLDYSIALYDGP
PVSYELPRIDPVDLPAITAPQVSGPLVFGNSGVVAVPVVFPMPVHRKQDAHRAE
PPVAAQGRRRSSSDVSAFQNEGSDDDSCSVQSAHNFHCGGRGTAAQEGRR
ADVTYGVTEDSRYESKEFDVYEQVAVTKKRGRTCSRCKGRKWKESKESCIVCDA
RECSYCVLRAMSGPEGRKCTICIGQPIDESKRLGKGRSLSRLLSPLEVRQILKA
EKECAQNLRLPEQLIINGFLNPDEMALLSCQRPQKLPGRVYWDKESGLMKEGE

Query Match 26.3%; Score 60.2; DB 8; Length 150120;
Best Local Similarity 64.2%; Pred. No. 1.8e-07;
Matches 106; Conservative 0; Mismatches 58; Indels 1; Gaps 1;
QY 65 acaagaataggcttgctgattgctgtaataatgggcaaaacttgccactcaattgct 124
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50458 ACTAGAAATGGTCTGCTGGCTTCGGGTATGGGGCAGACCTTGCCCTCAACATTCGA 50517
QY 125 tgaagaaggtctcccaatcccggttaacacggaaccatttcgaaggtattggccata 184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50518 GAGAAAGGG-TTCCCTATCTGTCTACACAGGACGACTTCTAAGGTGTGAGACCGT 50576
QY 185 agacgaagcaaacaccgaaggaacaccttcaatttatggggaacaa 229
|| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50577 TCAGCGCGCAAGGTAGAGAGAAACCTTCTCTGTGACGGGTTCGA 50621
RESULT 6
AF061838 1686 bp mRNA PLN 08-JUN-2000
LOCUS zea mays strain W64A x A182E putative cytosolic 6-phosphogluconate
DEFINITION dehydrogenase (pdh2) mRNA, complete cds.
ACCESSION AF061838
VERSION AF061838.1 GI:3342801
KEYWORDS zea mays.
SOURCE zea mays.
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
1 (bases 1 to 1686)
/note="Regulation of the oxidative pentose phosphate pathway in
maize (Zea mays L.) root plastids: induction of 6-phosphogluconate
dehydrogenase activity, protein and transcript levels
Plant Sci. 134 (2): 129-140 (1998)"
2 (bases 1 to 1686)
/note="Regulation of the oxidative pentose phosphate pathway in
maize (Zea mays L.) root plastids: induction of 6-phosphogluconate
dehydrogenase activity, protein and transcript levels
Plant Sci. 134 (2): 129-140 (1998)"
REFERENCE
AUTHORS Redinbaugh, M.G. and Campbell, W.H.
TITLE Nitrate regulation of the oxidative pentose phosphate pathway in
maize (Zea mays L.) root plastids: induction of 6-phosphogluconate
dehydrogenase activity, protein and transcript levels
JOURNAL Plant Sci. 134 (2): 129-140 (1998)
AUTHORS Redinbaugh, M.G. and Bailey-Serres, J.
TITLE Characterization of cDNAs encoding cytosolic and a putative
plastidic 6-phosphogluconate dehydrogenase

Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J.,
Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H.,
Schwartz, J., Southwick, A., Thaver, A., Toriumi, M., Vaysberg, M.,
Yu, G., Davis, R., Federspiel, N., Theologis, A., and Ecker, J.

Direct Submission

Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
On Jul 16, 1999 this sequence version replaced gi:5030434.

FEATURES

Location/Qualifiers
1..111222
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
/clone="F22C12"
complement(join(979..3122,3735..5930,6862..6892,
6983..7068,7655..9286,9346..10019,10228..10279,
10673..10816,11047..13429,13585..13835,14366..14645))
/note="similar to maize transposon MudR mudra protein
isolog gl11946355"

source

/codon_start=1
/evidence=not_experimental
/product="F22C12.1"
/protein_id="AAF24584.1"
/db_xref="GI:6692119"
/translation="MKTGGLITFTVILLASLPTKATGAGSRSKRDRESCERSRI
QTCLVNVSGIKISTECCFLKQPCLDVTKTKIKTNVLSRLKSCVFLSRDDI
AGFLNATQITISENPNELILLYLYNINLDDYTVVFLGFKFIICWILQLR
CLVNVLEIKIDFSEGEVAASVWVHNKDLAKFTLLWIPCRFVLMAMEVPVL
CWNRGTEVSGVYDSTPRMTKVRKTELILLDOLYLLGLDVRNRVRFGR
PPASQPMQLHFLLPVNVNDSLEMTLVPKSPKIKNVLEYLVKSDGVVSAARS
PILANPGSSILKRMDDIPVLEKNGSNGWDEDEDICNGRGMDDKNSDVAQ
VNLQNNISLKDQLQDSSGYSKCLSLWLDHDLVGLCFKIDICGLKAVDWCS
FGORSCEADKEDETEPCARKCKWTELAARMEKHEILIEIKTCHTSCALEP
QDSYMEFADFEFVRIHTQSTFELTKWYIEKIGYELQTSVRIAKNEAIRKVEG
IIVDTNLCKYQWKLMAISAVDANFFLLAFATTELSTDSWRFLSGIERVQR
CGQFEDLPALMAAIIHSSNGLLVWDYDLFPNPKFASFRGVAFWSQSTEGHRCPL
KGLCLSSPDPLLVINESGSOQEPWYNRECLRLHLSOFGRDYVLEDLVKRA
GTSQKEEDFSWKDIEKNSARKWLDQPPQWALANDHGRVYGIEMTEITFED
FNVSHLDNHLVLTGVLLDFELRHSDEFDEFFRGRGKCGNVYETPTEKLAERKDS
VTVDMPLDNAAFQVTAPOENDEMTVOLSDCTGEGFOSCKFPCLHAPLAKLLIN
PLEYVDTCYLERLYTYTATFSPVPELSAMPEASGPRLPVPIVPLPPPTNVAV
LITVCSAATVLMATREHMLLCTNACIKYGSQVYEGSSPKIRVKRNTENT
LISPTSVTLVMAETREHMLLCTNACIKYGSQVYEGSSPKIRVKRNTENT
LLNGLYIFGLAKQSKISIFGRYPVLPDLTYLHPFVNDSSLETLEVPNSVPS
IKVELYLEVSTSEGVDPAAACSPLENGLSSLRKQRTQBEAGYTHSVKLERVNG
LKVOGVNDSNGWIEDEADALNGSTHYVDREMTNKGSDGVEQVITLANNAESR
LCFDRDLKKTVDWCSIRGQKCVRETEKDYTFECIRKCKWLSQARATEHGLV
EITKNTPHTCPIGPDNDYVEFAADEIECLIRVQPTLIEELRNWFWNFGLMATS
EMQAKOEVIKYGWDQDFRVLPLNMAAFHPSNGLVVDWQYKLPNPEFASFSVF
WAFQSLTEGHCRCPLIVDTKOLGKYPMLKIASVEADDCYFFPLAFTAFVSSD
TWRFPSGIREKVTQKADICLSRPHDILDVINEGSOQEPWYNRECLDICTQF
HYVQDDYLANLVTEAGTSEKEDFIMNEIEKNSARKWLDQPPQWALANDHGRV
RYRMTIDAENLFAFESQSLGSPVATALLIDEMRFFYSGLDSSGSRNVRGM
YTPQVMPQLKMTSDISPHVMPLEKLFQVTEPQDEWIVQLSEWSTCGEFLFP
PCHLVLAQCEKILNQLVDCYSLDRLYKYAATFSPVEVAAMPEASGPTLFP
PVILPPNPGSKOWKTPRTINDKAKVPMDEELNAIVDLKVDLKTALADVLK
RLVNPVFCWNGIKDGTNGIYEGSNVPMVTKGTKEFNEILLHLYOVTGLDRKRS
KFGILRPDIQAFKVKRPLPVDDTSLTLEITTKHPYIKDIDLYLEVESTVG
VIDPATCSLWLDGSSGKOKTDIHWKLEREDRNSGQVHVLVDSTLNDKDSISG
HVKPCMSLWLDHDLRVGLCFKQDELKAVDMCSLKAQKCVKRETKADYIEFC
IRWCKWSLGAARKKHGLVEIKYTGHTCHPIVPEDFSEFETIEIRAVRYMPTQ
TISELKKWKKIKGYELTSDVRLAKEKAKIRVGDWQDFEDFPKLMSALCNSGILL
VDWYDLFPNPFASFCGVFAWQPSLEGHCRCPLIVVDTKNLCYQLKLIASGV
QWQEPWAYHFLNHFYSQSFVPSFCGLGARIRAGSTSQKDEFFSYMNDIEKKEPE
ARKWLDQPPQWALANDHGRVYIMETNKALFAVCAFEAGHVYVGLVLLDEL
RSKDKFSQSRSLNGDGVYTEPMDKLEEFRTFTYIYITVDDNNAFQVATADL
KQECIVOLSDCSTCGDFQRYKPCPLHALAVCKIKENFLPQVDDCYTLERLKRAT
IFSHVPMASWPEASGPRLLPVPVPPSPPTVYVSRVQYKVMCHLNGYIKYQAD
GVYEGSTLTKITIMRTALSRLLDRLYQLFGLDKQSEFKIFGKYPVDSPLFTVY
HFPVNDSSLETLEVPNSVNLFPYLEAQPTSDAVVVPVTCSSPLESPDSAKR
QRITQOEIVDNNSSAGILVVKVNSAGLPCLLPRLWIDDDHMHGLCLCFKDRDELK

CDS

/codon_start=1
/evidence=not_experimental
/product="F22C12.2"
/protein_id="AAF24583.1"
/db_xref="GI:6692118"
/translation="MASKRIILKELKQKDPPTSCSAVAEDMFHWQATIMGPSDSPYS
GGFLVTHFPPDPYFPKPKVAFKVFHPNVNSNGSICLDILKEQWSPALTISKVLL
SICLLIDPNDDPLVPEIAHMYKTDRAKYESTARSWTQYAMG"
complement(join(18031..18617,18685..19861))
/note="putative receptor kinase emb|CAA3040.1"

CDS

/codon_start=1
/evidence=not_experimental
/product="F22C12.3"
/protein_id="AAF24582.1"
/db_xref="GI:6692117"

/translation="MQIFLFFSLICFVLISSTOLEDKALLHLFSSNSRLHWN
QSSDVCHSWTNCENGDRIVSVRLPAVGNGLIPPTISRLSLKFLSKRNHFTG
DPFSDNLKSLTHLYLQHNHLSGPLLAIFSELKNLKLVDLSNNNGSIPTSGLT
SLOVLNANNFSGEIPNLHLPKLSQINLNNKLGITPKSLQRFQSFSSNNLTER
KKQRTPLGSLQALFLILSAACVLCVGLSFMITCFGKTRISGKLRKRSPPGPN
KTRSDNVTEEGKIFFGGRNHLFDLDDLSSSEVLGKGAFTTYKVTMEDMTVVV
KRLVGVVREFQEQIEMIRHENVAELKAIYYSKDDKLAVSYNNGSLPEILH
GNRGYRVPDLDMDARIATGAARGLAKTHEGFIHGNKSSNIFLDSQCYGIGDV
GLTTIMRSLPOTTCITSGYHAPETDTRSTQFSDVYSGVYLLELTGKSPVSAQ
VPTGGENMDLASWIRSVVAKWGEVDFMELTSQGGFEEMVEMQLGLACVALKQ
BRPHTAOLVLIEDIRSVDAF"

CDS

/join(20892..20993,21210..21310,21520..21694,21780..21976,
22063..22162,22282..22387,23017..23183)
/note="similar to vacuolar ATPase gi|3600058"

/codon_start=1
/evidence=not_experimental
/product="F22C12.4"
/db_xref="GI:6692094"
/protein_id="AAF24559.1"

/translation="MNDADASTIQOOWVRFTQRAEAEKANEISISSEEFNLEKQLV
EAEKKIRQYEYKKEKQVDVRKKIDYSQMLNARITKVLQAOQDIDVNAKMEAAKOLK
VSOHGFNNHHQYKHLKLDLIVQCLRLKEPALLRCREEDLDIVESMLDDASEYC
KAKYHAPETIVDKDIFLPAPSDDDPHALSCAGGVILASRDGKIVCENTLDARLEVA
FENKLPFCSSGSLFEMCMVDKVALRCQWCSLMDSNFITTEKLRDAKSMNPTGRCP
DNGVEKSSMCYSKCTQGMGSGSQGHKGNVMEBCYEG"

CDS

/note="putative 6-phosphogluconate dehydrogenase
dbj|BAA22812; similar to Ests emb|F15308, emb|I15307,
gb|I21894, gb|A1997770.1"
/codon_start=1
/evidence=not_experimental
/product="F22C12.5"

/db_xref="GI:6692095"
/protein_id="AAF24560.1"
/translation="MESALISRIGLAGLAVGMQNLALNAEKGFPISVYNTTTSKVD
TIDRAVEGNLPSGOYSPRDFVLSIQPRSLIILVAGAPVDQIDAFSVMPEGDC
TIDGNEWQNTERRISEAQGLVYLGWGGEGEAGNGSLPMGFSQAFQADNKK
ILEKYAAQVEDGPCVTYIEGSGNGFVKNHNGIEYQMOLISEAYDVLKNVGGLSNE
ELAEIPTENWSELESFIVLEITSDIRKVDGDELDVLDKTKMGKCTKWTYQQA
AELSVAKTCISYAAQGNLIRAKSLSEKSNLNFEGELRIWKGCCIIRAVFLDRKRAKL
RNPDLASLVDPDEFAEMVQQAARVVRVGLAVSGAISTPGMCASLAYFDTRYRARLP
ANLVQAQRDLFGAHTYERTDRPGAYHTIEWTKLARKNH"

Sat Nov 4 18:11:16 2000

/clone_lib="RPCI-11 human BAC library"

/map="17"
/chromosome="17"
2. .110
/rpt_family="L2"
399. .427
/rpt_family="AT-rich"
complement(593. .1090)
/rpt_family="LMB7"
1477. .1635
/rpt_family="MIR"
2171. .2861
/rpt_family="LIPAL0"
complement(2899. .3103)
/rpt_family="MIR"
3384. .3442
/rpt_family="G-rich"
4482. .4564
/rpt_family="MIR"
complement(4565. .4746)
/rpt_family="MER69A"
4747. .4784
/rpt_family="MIR"
4788. .4966
/rpt_family="(TTTTC)n"
complement(4967. .5259)
/rpt_family="ALUSx"
5260. .5357
/rpt_family="MIR"
5362. .5418
/rpt_family="(CA)n"
complement(5523. .5832)
/rpt_family="ALUSx"
7567. .7732
/rpt_family="MER63A"
7751. .7882
/rpt_family="MIR"
8008. .8286
/rpt_family="L2"
9311. .9531
/rpt_family="MIR"
complement(9756. .9890)
/rpt_family="L2"
complement(10044. .10344)
/rpt_family="ALUSp"
complement(10845. .11308)
/rpt_family="L2"
complement(11335. .11458)
/rpt_family="MER69A"
complement(11619. .11674)
/rpt_family="MER69A"
complement(11680. .12175)
/rpt_family="L2"
12176. .12204
/rpt_family="(TTTTTA)n"
complement(12206. .12454)
/rpt_family="AluJb"
complement(12455. .12843)
/rpt_family="L2"
13247. .13529
/rpt_family="ALUSx"
complement(14301. .14370)
/rpt_family="MIR"
15962. .16265
/rpt_family="ALUSg"
complement(16706. .16785)
/rpt_family="MIR"
complement(16804. .16875)
/rpt_family="L2"
18053. .18195
/rpt_family="HAL1"
18328. .19612
/rpt_family="HAL1"

repeat_region complement(19638. .19830)
/rpt_family="MIR"
19857. .19907
/rpt_family="A-rich"
19951. .20217
/rpt_family="MLTII"
complement(20912. .21095)
/rpt_family="MIR"
21171. .21213
/rpt_family="AT-rich"
21390. .21534
/rpt_family="MER103"
complement(22377. .22503)
/rpt_family="AluJo/FRAM"
22877. .22973
/rpt_family="MIR"
23260. .23297
/rpt_family="GA-rich"
24711. .24949
/rpt_family="MIR"
25364. .25407
/rpt_family="MIR"
complement(25408. .25718)
/rpt_family="ALUSx"
complement(25731. .26012)
/rpt_family="ALUSx"
26919. .26946
/rpt_family="AT-rich"
27919. .28060
/rpt_family="MIR"
complement(29274. .29446)
/rpt_family="FRAM"
complement(29447. .29548)
/rpt_family="MIR"
complement(30115. .30239)
/rpt_family="HAL1"
complement(30359. .30444)
/rpt_family="LIME"
32686. .32727
/rpt_family="(TG)n"
33941. .34089
/rpt_family="FRAM"
34120. .34271
/rpt_family="L2"
34650. .34683
/rpt_family="(CA)n"
34868. .34897
/rpt_family="(TATATG)n"
complement(35799. .36115)
/rpt_family="ALUSq"
37092. .37423
/rpt_family="L2"
38231. .38278
/rpt_family="CT-rich"
complement(38610. .38723)
/rpt_family="ALUSg/x"
complement(38749. .38823)
/rpt_family="MER63C"
complement(38824. .39178)
/rpt_family="LIPB1"
39180. .39214
/rpt_family="(TA)n"
complement(39215. .39605)
/rpt_family="MER63C"
39618. .39667
/rpt_family="(TA)n"
40234. .40269

Query Match 16.7%; Score 38.2; DB 10; Length 170453;
Best Local Similarity 53.7%; Pred. No. 0.64;
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 68 agaataggccttgctgattgctgttaattgagcaaatctggcaaatctgctcaattgttga 127

```

Db 1879 AAAAGAGGCTCTCTCTATTATCTGTGAAGAGCAAAATCTCTCACCTTGCCCTGG 1820
QY 128 aaagggtcccaattccggttaacaaaggaacatttccaggttatgggccaataaga 187
Db 1819 GAAGGAAAAGAGAGTATTGTGTGACAGAGTCTCTATGAGCAAGTAATGTCGAATGAGA 1760
QY 188 cgaagcaaacagggaaggaacattca 214
Db 1759 GTATTTCACAGAAAGGCCCTCTCA 1733

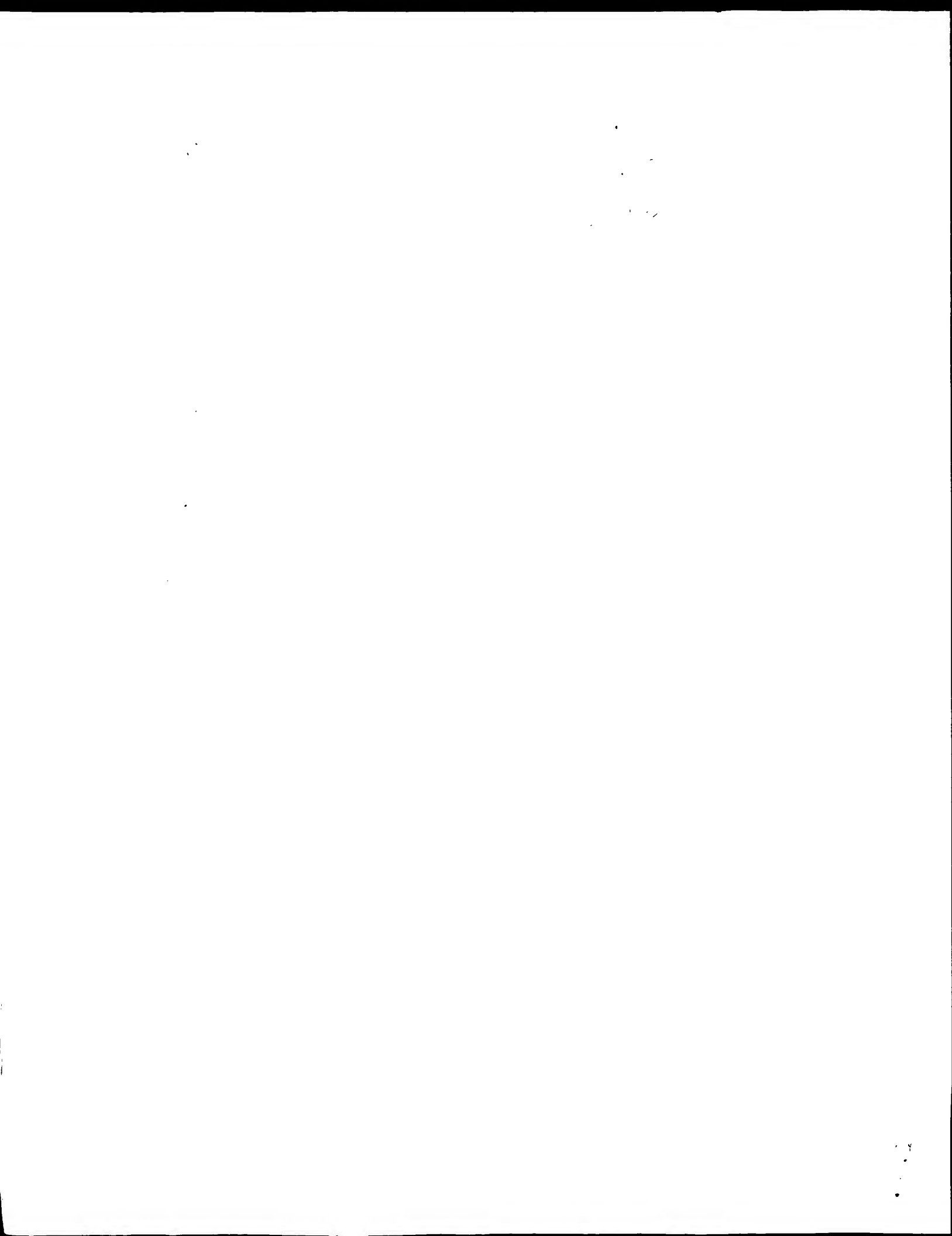
RESULT 13
AC015724/c
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone RP11-45J24 map 17, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC015724
AC015724.4 GI:7329270
VERSION
KEYWORDS
SOURCE
HTG: HTGS_PHASE1; HTGS_DRAFT.
ORGANISM - Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184799)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,K., Gage,D.,
Gaigani,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGuck,K., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,K., Roy,A., Santos,R., Severi,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 26, 2000 this sequence version replaced gi:6751809.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1414
Center clone name: 45_J24
----- Summary Statistics
Sequencing vector: M13; M7815; 96% of reads
Sequencing vector: Plasmid; n/a; 8-0.1% of reads
3.81918146119Chemistry: Dye-terminator Big Dye; 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 164567 bases at least Q40
Consensus quality: 174814 bases at least Q30
Consensus quality: 179219 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 18319; sum-of-ctigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality cov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces

```

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2143: contig of 2143 bp in length
* 2144 2243: gap of 100 bp
* 2244 6178: contig of 3935 bp in length
* 6179 6278: gap of 100 bp
* 6279 10043: contig of 3765 bp in length
* 10044 10143: gap of 100 bp
* 10144 12960: contig of 2817 bp in length
* 12961 13060: gap of 100 bp
* 13061 17800: contig of 4740 bp in length
* 17801 17900: gap of 100 bp
* 17901 22268: contig of 4368 bp in length
* 22269 22368: gap of 100 bp
* 22369 27281: contig of 4913 bp in length
* 27282 27381: gap of 100 bp
* 27382 32880: contig of 5499 bp in length
* 32881 32980: gap of 100 bp
* 32981 41040: contig of 8060 bp in length
* 41041 41140: gap of 100 bp
* 41141 48535: contig of 7395 bp in length
* 48536 48635: gap of 100 bp
* 48636 56130: contig of 7495 bp in length
* 56131 56230: gap of 100 bp
* 56231 71938: contig of 15708 bp in length
* 71939 72038: gap of 100 bp
* 72039 86256: contig of 14218 bp in length
* 86257 86356: gap of 100 bp
* 86357 100509: contig of 14153 bp in length
* 100510 100609: gap of 100 bp
* 100610 122739: contig of 22130 bp in length
* 122740 122839: gap of 100 bp
* 122840 144885: contig of 22046 bp in length
* 144886 144985: gap of 100 bp
* 144986 184799: contig of 39814 bp in length.
FEATURES
source
1. 184799
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone_lib="RP11-45J24"
/clone="RP11-45J24"
1. 2143
/note="assembly_fragment"
2244. 6178
/note="assembly_fragment"
6279. 10043
/note="assembly_fragment"
10144. 12960
/note="assembly_fragment"
13061. 17800
/note="assembly_fragment"
17901. 22268
/note="assembly_fragment"
22369. 27281
/note="assembly_fragment"
27382. 32880
/note="assembly_fragment"
clone_end:SP6
vector_side:right
32981. 41040
/note="assembly_fragment"
clone_end:T7
vector_side:right
41141. 48535
/note="assembly_fragment"
48636. 56130
/note="assembly_fragment"

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:01 ; Search time 4352.3 Seconds
(without alignments)
582.439 Million cell updates/sec

Title: US-09-300-482-14
Perfect score: 410
Sequence: 1 cccacgcgtccgggtcatg.....gcaacggccgcctcttgatg 410

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estba.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estom.*
68: em_estov1.*
69: em_estov2.*
70: em_estpl1.*
71: em_estpl2.*
72: em_estpl3.*
73: em_estpl4.*
74: em_estpl5.*
75: em_estro1.*
76: em_estro2.*
77: em_estro3.*
78: em_estro4.*
79: em_estro5.*
80: em_estro6.*
81: em_estro7.*
82: em_estro8.*
83: em_estro9.*
84: em_estro10.*
85: em_estro11.*
86: em_estro12.*
87: em_estro13.*
88: gb_gss1.*
89: gb_gss2.*
90: gb_gss3.*
91: gb_gss4.*
92: em_gss1.*
93: em_gss2.*
94: em_gss3.*
95: em_gss4.*
96: gb_gss5.*
97: gb_gss6.*
98: gb_gss7.*
99: gb_gss8.*
100: gb_gss9.*
101: em_gss5.*
102: em_gss6.*
103: em_gss7.*
104: em_gss8.*
105: em_gss9.*
106: em_gss10.*
107: em_gss11.*
108: gb_gss10.*
109: gb_gss11.*
110: em_gss12.*
111: gb_gss12.*
112: gb_gss13.*
113: gb_gss14.*
114: gb_gss15.*
115: gb_gss16.*
116: gb_gss17.*

pBluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside.
Plasmid DNA Preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT 121 a 149 c 160 g 117 t
ORIGIN

Query Match 77.8%; Score 319; DB 36; Length 547;
Best Local Similarity 87.2%; Pred. No. 1.8e-73;
Matches 349; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 11 cgcggtcaggggagaaacctgcctcaacattgcagagaaaggttccccatctctgt 70
Db 48 CCGGTCATGGGGCAGAACCTTGCCTCAACATTGCCGAGAAAGCTTCCCATCTCTGT 107

QY 71 gtacaacaggacaacctccaaggtgagacagcgtgcagcgtgccaaaggcagaagaaa 130
Db 108 CTACAACAGGACACAGCTCCAAGGTGGATGAGACTGTTTCAGCGTGCCAAAGCTAGAAGGAAA 167

QY 131 ccttcctctacggtctccatgaccccgctcctttgtgaagtcattcagaagccagc 190
Db 168 CTTCTCTCTATGTTTCCATGACCCCTGCATCTCTTGTCAACTCCATTCAGAAAGCCACG 227

QY 191 ggtgtgatcatgctcgtcaaggccgcgcgcagttgacacagaccatcgacgctcgc 250
Db 228 CGTCGTCATCATGCTTGTCAGAGGTGGTGCTCCGGTTGACACAGCATCGGACGCTCGC 287

QY 251 agtcaacttgagcagggcgactgcacatgcattgagggggaacagtggttacgagaacac 310
Db 288 TGCACACCTGGAGCAAGGTGACTGTCATGTTGATGGAGAAACAGTGATGAGAACAC 347

QY 311 ggagaggaggaagccatggagagcgcgccctnctgtatcttggcattgggtgtctc 370
Db 348 CGAAGGAGGAGGAGGAGGCGATGGAGGAGCGTGAGCTCTTACCTCGGATGGGTGTTT 407

QY 371 tggaggaaaggaggtgcccccaacgcccgcctctctgatg 410
Db 408 CGGAGGAGGAGGAGGCGCGCCCGCATGGCCGCTCCATGATG 447

RESULT 6
AW563648 519 bp mRNA EST 19-JUL-2000
LOCUS LG1_245_A11.b1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
ACCESSION AW563648
VERSION AW563648.1 GI:7217526
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
1 (bases 1 to 519)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
An EST database from Sorghum: light-grown seedlings
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 518
POLYA-No. Location/Qualifiers
1..519
/organism="Sorghum bicolor"

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LG1)"
/notes="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI
; The library was made from poly-A RNA in the cloning
vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT 118 a 137 c 157 g 107 t
ORIGIN

Query Match 76.7%; Score 314.6; DB 23; Length 519;
Best Local Similarity 91.7%; Pred. No. 2.5e-72;
Matches 332; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 12 gcggtcatggggcagaaacctgcctcaacattgcagagaaaggttccccatctctgt 71
Db 158 GCGGTCATGGGGCAGAACCTTGCCTCAACATTGCAGAAAGGCTTCCCATCTCTGTC 217

QY 72 tacaacaggacaacctccaaggtgagacagcgtgcagcgtgccaaaggcagaagaaaac 131
Db 218 TACAACAGGACAACCTCCAAGGTTGATGAGACCGTGCAGCGTGCCAAAGGCGAGAAGAAAC 277

QY 132 cttccgtctacggtctccatgaccccgctcctttgtgaagtcattcagaagccagc 191
Db 278 CTTCCGTGTGTACGGCTTCCATGACCTGCATCCTTTGTGAACCTCCATTCAGAAGCCGCT 337

QY 192 gttgtgatcatgctcgtcaaggccgcgcgcagttgacacagaccatcgacgctcgc 251
Db 338 GTTGTATCATGCTTGTCAAGGCTGGTGACCACTGACACAGCATTCGAGCGCTCGCA 397

QY 252 gctcaacttgagcagggcgactgcacatgcattgagggggaacagtggttacgagaacac 311
Db 398 GCTCACTTGGAGCAGGCTGACTGATCATTTGATGGAGGAATGAGTGGTATGAGAACA 457

QY 312 gagagaggaagagccatggagagcgcgccctnctgtatcttggcattgggtgtctc 371
Db 458 GAGAGGAGGAGAAAGGCCATGGAGGAGCGTGGCCCTCTATCTTTGGAATGGGCGCTCT 517

QY 372 gg 373
Db 518 GG 519

RESULT 7
BE363025 516 bp mRNA EST 20-JUL-2000
LOCUS DGL_9_A10.b1_A002 Dark Grown 1 (DGL) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.
ACCESSION BE363025
VERSION BE363025.1 GI:9304582
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
1 (bases 1 to 516)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.
An EST database from Sorghum: dark-grown seedlings
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 516

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

7

Fax: 0298-38-7468
Email: minobes@rics0.riken.go.jp
PROJECT = "RGP"

FEATURES

source
Location/Qualifiers
1. .456
/organism="Oryza sativa"
/strain="Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="R2869_1A"
/clone_lib="Rice root"
/note="Prepared from seedling root."
107 a 104 c 130 g 112 t 3 others

BASE COUNT 107 a 104 c 130 g 112 t 3 others

Query Match 70.8%; Score 290.4; DB 36; Length 456;

Best Local Similarity 85.6%; Pred. No. 5.2e-66;

Matches 332; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

QY 12 gcggctatggggcagaaaccttgcctcaacattgcagagaagggttccccatctctgtg 71

DB 70 GCGGTCATGGGCGAGAACCTTGCCCTCAACATTCAGAGAAAGGTTCCCTATCTCTGTC 129

QY 72 tacaacaggaacaacctccaagtggaagcagaccgtgacggtgcagagcagagaaggaac 131

DB 130 TACAACAGGACGACTTCTAAGGTTGATGAGACCGTTCAGCGCGCCAGGTAGAGGAAC 189

QY 132 cttcccgctacggtccatccatgaccccgctccttcttgagtccattcagaagccacg 191

DB 190 CTTCTGTGTACGGGTTTCATGACCCCTGCATNCTTTGTGAACCTCAATTCAGAACCCAGT 249

QY 192 gtggtgatcatgctgcagagccgagcgcgcaggtgacagaccatcgagagcgtcgca 251

DB 250 GTTGTATCATGCTGTCAAGGCTGGTGACCACTGAGGACGAGACCATTCGAACTCTTGCA 309

QY 252 gctcaactggagcggcgagcagtcgacatcgatggggggaacgagtggtgacgagaacacg 311

DB 310 GCACACTTGGAGCAGGCTGACTGATTATTGATGAGGAAATGATGTTGACGAGACACT 369

QY 312 gagagaggagagagccatgagagagcggcgccctctgtatcttggcagtggtgtctct 371

DB 370 GAGAGGAGGAGAGGAAGAAATGAGAGCGTTGNCCTCTATCTATCTTNGATGGTGTTC 429

QY 372 ggaggaaggagggtgcccgaacgccc 399

DB 430 GGA-GAGAGGAGGTTCCCGAATTGGCC 456

RESULT 10

BE497862

LOCUS

DEFINITION WHE0957_D04_H072S Wheat pre-anthesis spike cDNA library Triticum aestivum CDNA clone WHE0957_D04_H07, mRNA sequence.

ACCESSION BE497862

VERSION BE497862.1

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Triticum.

AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.

TITLE The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library

JOURNAL Unpublished (2000)

COMMENT Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oandersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20

Seq primer: Stratagene SK primer.

FEATURES

source

Location/Qualifiers

1. .472

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db_xref="taxon:4565"

/clone="WHE0957_D04_H07"

/clone_lib="Wheat pre-anthesis spike cDNA library"

/tissue_type="Spike before anthesis"

/dev_stage="Adult plant"

/lab_host="E. coli SOLR"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: ECOLI; Site 2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phagescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

110 a 145 c 124 g 93 t

BASE COUNT 110 a 145 c 124 g 93 t

ORIGIN

Query Match 70.8%; Score 290.2; DB 36; Length 472;

Best Local Similarity 87.8%; Pred. No. 5.9e-66;

Matches 316; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 11 cgcggtcatggggcagaaaccttgcctcaacattgcagagaagggttccccatctctgt 70

DB 112 GCGGTCATGGGCGAGAACCTTGCCCTCAACATTCGAGAGAAAGGCTTCCCATCTCTGT 171

QY 71 gtacaacaggaacaacctccaagtggaagcagcgtgcagcgtgcagagcgagaaggaac 130

DB 172 CTACAACAGGACGACCTCCCAAGGTCGATGAGACTGTCCAGCGCGCAAGCTAGAGGAAA 231

QY 131 cttcccgctacggtctccatgaccccgctccttcttgagtccattcagaagccacg 190

DB 232 CTTTCCCTGTAGGTTTCCATGACCCCTCATCTCTGTCACACTCCATTCAGAACCCAG 291

QY 191 ggtggtgatcatgctgcagagcggcgccaggtgaccagacacatcgagcgtcgc 250

DB 292 CGTGTGTCATCATGCTTGTCAAGGCTGGCGCACCCAGTTGACGACGACCATCGCCACGCTCGC 351

QY 251 agctcaactggagcagggcgactgcatcatcgatggggggaacgagtggtgacgagaacac 310

DB 352 AGCACACCTGGAGCAGGGCGGACTGTCATCGTTGATGAGAGAAACGAGTGTATGAGAACAC 411

QY 311 ggagagggagggagggccatggagagcgccctctctgtatcttgcagtggtgtctc 370

DB 412 AGAAGAGGAGGAGAGGCCATGGAGGCGGTGACTCTCTACCTCGGATGGGTGTTTC 471

RESULT 11

BE497880

LOCUS

DEFINITION WHE0957_B09_D172S Wheat pre-anthesis spike cDNA library Triticum aestivum CDNA clone WHE0957_B09_D17, mRNA sequence;

ACCESSION BE497880

VERSION BE497880.1

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.

AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.

TITLE The structure and function of the expressed portion of the wheat
JOURNAL genomes - Pre-anthesis spike cDNA library
COMMENT Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Email: oandersn@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.

FEATURES

source
 1. .474
 Location/Qualifiers
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone_lib="WHE0957_B09_D17"
 /clone_lib="Wheat pre-anthesis spike cDNA library"
 /tissue_type="Spike before anthesis"
 /dev_stage="Adult plant"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
 greenhouse. Whole spike with awns trimmed, white, green
 and yellow anther were collected and total RNA, and
 poly(A) RNA were prepared, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 phagemids in the TJ Close lab (Choi, Close, Fenton) at
 the University of California, Riverside. Plasmid DNA
 preparations and DNA sequencing were performed in the OD
 Anderson lab (all other authors)."
 BASE COUNT 107 a 143 c 130 g 94 t

Query Match 70.4%; Score 288.6; DB 36; Length 474;
 Best Local Similarity 87.5%; Pred. No. 1.5e-65;
 Matches 315; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
 QY 11 cgcggtcatggggcagaacctgcctcaacattgcagagaaaggttcccatctctgt 70
 Db 115 CGCCGTCATGGGCGAGAACCTCGCCCTCAACATTTGGGAGAAAGGTTCCCATCTCTGT 174
 QY 71 gtacaacaggacaacctccaagggtgagacgagccgtgcaggtgccaaggcagaagaaa 130
 Db 175 CTACAACAGGACCACTCCCAAGGTGATGAGACTGTTTCAGCGCGCCCAAGCTAGAGGAAA 234
 QY 131 ctttcccgctctacggtcttcattgaccccgcgctcttttgaagtcattcagaagccacg 190
 Db 235 CTTTCTCTTACGGTTTCCATGACCCCTGCATCTCTGTCACCTCCATTCAAGAGCCACG 294
 QY 191 gdtggtgatcatgctcgtcaaggcggcgccagtgccagtgaccagaccatcgcgagctcgc 250
 Db 295 TGTGTCATCATGCTGTGTCAAGGCGGTTGCCCGGTTGACCCAGACCATTTCAAGCGTCGC 354
 QY 251 agctcaactggagcaggcgactgcatcatgatggggggaacagtggtgacagaacac 310
 Db 355 AGCACACCTGGAGAGGCGACTGTCATGTTGATGAGGGAACGAGTGGTATGAGAACAC 414
 QY 311 ggagaggaggagagggcagcagtgagagcgcgccctnctgtatcttggcatgggtgtctc 370
 Db 415 GGAAGGAGGAGAGGCGATGGAGGAGCGTGGACTCTCTACCTCGGGATGGGTGTTTC 474

RESULT 12
 BE425231 512 bp mRNA EST 24-JUL-2000
 LOCUS WHE0312_G04_G0425 Wheat unstressed seedling shoot cDNA library
 DEFINITION Triticum aestivum cDNA clone WHE0312_G04_G04, mRNA sequence.
 ACCESSION BE425231
 VERSION BE425231.1 GI:9423146

KEYWORDS

bread wheat.
 Triticum aestivum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
 1 (bases 1 to 512)
 Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
 P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
 Seaton, C.B. and Tong, J.C.
 The structure and function of the expressed portion of the wheat
 genomes

TITLE

Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oandersn@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stratagene SK primer.

FEATURES

source
 1. .512
 Location/Qualifiers
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone_lib="WHE0312_G04_G04"
 /clone_lib="Wheat unstressed seedling shoot cDNA library"
 /tissue_type="Etiolated shoot"
 /dev_stage="Five day old seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
 , germinated and grown aseptically in the dark at room
 temperature on filter paper with water, nystatin and
 cefotaxime in covered crystallization dishes. Shoots were
 harvested. The tissue, total RNA, and poly(A) RNA clones
 prepared, a cDNA library was made, and the cDNA clones
 were in vivo excised to give pBluescript phagemids in the
 TJ Close lab (Choi, Close, Fenton) at the University of
 California, Riverside. Plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."
 BASE COUNT 116 a 142 c 144 g 110 t

Query Match 70.4%; Score 288.6; DB 35; Length 512;
 Best Local Similarity 85.8%; Pred. No. 1.6e-65;
 Matches 343; Conservative 0; Mismatches 55; Indels 2; Gaps 2;
 QY 11 cgcggtcatggggcagaacctgcctcaacattgcagagaaaggttcccatctctgt 70
 Db 15 CGCGGTCATGGGCGCAACCTTGCCCTC-ACATTCGCGAGAAA-GCTTCCCATCTCTGT 72
 QY 71 gtacaacaggacaacctccaagggtgagacgacctgacgagcagtgccagtgccaaggcagaagaaa 130
 Db 73 CTACAACAGGACCACTCCCAAGGTGATGAGACTGTTTACAGCTGCGCAAGCTAGAGGAAA 132
 QY 131 ccttcccgctctacggtcttcattgaccccgctcttttgaagtcattcagaagccacg 190
 Db 133 CCTTCTCTATATGTTTTCATGACCTTCATCTTTTGTCAACTCCATTCAGAACGACG 192
 QY 191 gdtggtgatcatgctcgtcaaggcggcgccagtgccagtgaccagaccatcgcgagctcgc 250
 Db 193 TGTGTCATCATGCTGTGTCAAGGCTGGTGTCTCGGTGACGACCATTCGACGCTCGC 252
 QY 251 agctcaactggagcaggcgactgcatcatcgatggggggaacagtggtacagaacac 310
 Db 253 AGCACACCTGGAGCAAGGTGACTGTCATCGTTGACGAGGAGAAACGAGTGGTATGAGAACAC 312
 QY 311 ggagaggaggagagggcagcagtgagagcgcgccctnctgtatcttggcatgggtgtctc 370

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:06 ; Search time 4352.3 Seconds
(without alignments)
448.904 Million cell updates/sec

Title: US-09-300-482-225
Perfect score: 316

Sequence: 1 gataaggtgcgcacactgag.....gaatggaggaggactacttoga 316

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estrol:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

LOCUS	AU088734	575 bp	mRNA	EST	31-MAR-2000
DEFINITION	Rice shoot Oryza sativa cDNA clone s4976, mRNA sequence.				
ACCESSION	AU088734				
VERSION	AU088734.1	GI:7378463			
KEYWORDS	EST.				
SOURCE	Oryza sativa.				
ORGANISM	Oryza sativa				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.				
AUTHORS	1 (bases 1 to 575)				
TITLE	Sasaki, T. and Yamamoto, K.				
JOURNAL	Rice cDNA from etiolated shoot (2000)				
COMMENT	Unpublished (2000)				
	Contact: Takuji Sasaki				
	National Institute of Agrobiological Resources				
	Rice Genome Research Program				
	2-1-2 Kannondai, Tsukuba				
	Ibaraki,				
	Japan 305				
	Tel: 0298-38-7441				
	Fax: 0298-38-7468				
	Email: tsasaki@agr.affrc.go.jp				
	PROJECT = 'RGP'.				
	S4976-8Z.				

```

FEATURES
  source
    Location/Qualifiers
      1..575
        /organism="Oryza sativa"
        /strain="Nipponbare, sub_species Japonica."
        /db_xref="taxon:4530"
        /clone="S4976"
        /clone_lib="Rice shoot"
        /note="Etiolated shoot (8 days old)"
BASE COUNT      153 a      109 c      153 g      160 t
ORIGIN
      Query Match      50.9%;      Score 160.8;      DB 15;      Length 575;
      Best Local Similarity      83.2%;      pred. No. 1.3e-39;
      Sub-192. Conservative      0;      Mismatches 37;      Indels 0;      Gaps 0;

```

[illegible]

RESULT	2	
LOCUS	AU031584	
DEFINITION	AU031584 Rice cDNA from immature leaf including apical meristem	EST 30-OCT-1998
ACCESSION	AU031584	mRNA
VERSION	AU031584.1	GI:3767474
KEYWORDS	EST.	
SOURCE	Oryza sativa.	
ORGANISM	Oryza sativa.	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta	
AUTHORS	Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.	
TITLE	1 (Bases 1 to 369)	
JOURNAL	Sasaki, T. and Yamamoto, K.	
	Rice cDNA from immature leaf including apical meristem	
	Unpublished (1997)	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	160.8	50.9	575	15	AU088734	AU088734 AU088734
2	158.8	50.3	369	15	AU031584	AU031584 AU031584
3	158.8	50.3	403	36	D22889	D22889 RICC1720A R
4	150.2	47.5	499	36	C98335	C98335 C98335 Rice
5	124.4	39.4	472	36	C73011	C73011 C73011 Rice
6	118.4	37.5	504	14	AL380932	AL380932 MPEB55E04
7	118.4	37.5	542	21	AL380933	AL380933 MCB555E04
8	116	36.7	638	21	AW350186	AW350186 GMB210007B
9	116	36.7	736	21	AW348508	AW348508 GM210002B
10	106.6	33.7	165	35	BE359923	BE359923 DGL_58_D0
11	100.8	31.9	749	33	BE034126	BE034126 MG05C09 M
12	97.2	30.8	483	14	AI993279	AI993279 701495920
13	95	30.1	544	19	AV545174	AV545174 AV545174
14	94.6	29.9	619	39	AI999119	AI999119 701554529
15	90.2	28.5	317	39	T75815	T75815 10593 Lambda
16	85.8	27.2	257	21	AT289686	AT289686 NNXV004D0
17	85.2	27.0	510	14	AL371121	AL371121 MCB4A2A06
18	77.8	24.6	653	13	AI896020	AI896020 EST265463
19	71.2	22.5	635	21	AT288562	AT288562 618067F10
20	70.6	22.3	556	12	AI745746	AI745746 605076F04
21	68.2	21.6	646	14	AI947960	AI947960 603031G04
22	67.2	21.3	605	11	AI622577	AI622577 486058F10
23	66.6	21.1	562	12	AI711616	AI711616 605058C06
24	65.4	20.7	525	111	A0844969	A0844969 an41c06_3
25	63.8	20.2	245	25	AW927596	AW927596 945006E06
26	63.8	20.2	475	25	AW928010	AW928010 945006E06
27	63	19.9	485	19	AW017710	AW017710 614062C03
28	63	19.9	523	13	AW065819	AW065819 614062C03
29	63	19.9	535	13	AI783104	AI783104 614008E11
30	63	19.9	565	13	AI783103	AI783103 614008E10
31	63	19.9	593	19	AW000034	AW000034 614008E10
32	62	19.6	562	33	BE056891	BE056891 945034F03
33	60.8	19.2	378	14	AI943885	AI943885 614031A12
34	60.8	19.2	414	22	AW498391	AW498391 660047C09
35	60.8	19.2	460	19	AW052901	AW052901 614075F10
36	60.8	19.2	490	36	BE518996	BE518996 946075E06
37	59.8	18.9	515	22	AW438351	AW438351 707071B11
38	58.2	18.4	495	14	AW091086	AW091086 614090A07
39	58	18.4	825	25	AW983286	AW983286 HVSME9001
40	57.8	18.3	424	40	W80360	W80360 zh49h03_s1
41	56.8	18.0	448	13	AI891201	AI891201 614020B05
42	56.4	17.8	325	9	AI168746	AI168746 ox66802_s
43	55.6	17.6	490	20	AW191777	AW191777 687033B06
44	55	17.4	394	14	AI947921	AI947921 603031A12
45	55	17.4	613	5	AA556420	AA556420 275 Loblo

ALIGNMENTS

RESULT 1
- AH088734

COMMENT

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.

FEATURES

source
location/Qualifiers
1..369

/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E51890.2Z"
/clone_lib="Rice cDNA from immature leaf including apical meristem"
/dev_stage="immature"
/note="Organ: leaf; immature leaf including apical meristem (under long day condition)";

BASE COUNT 115 a 71 c 90 g 88 t 5 others
ORIGIN

Query Match 50.3%; Score 158.8; DB 15; Length 369;
Best Local Similarity 82.3%; Pred. No. 5e-39;
Matches 181; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 gataagtgccacactgagaagaagtaacctcccttgacatagaggttgatggtgt 60
Db 4 GAAAGGTGCGTCCCTTGAGAGAGAGTACCCATCCCTTGACATCGAGGTTGATGTTGT 63
Qy 61 ctaggctctcaacactagacgtgcccgcacatctgtcggggcccaattgcctgcctgga 120
Db 64 CTGGTCTTCCTCCACATGATGATGCGTGCATCTGCTGGGCTAACTGCTGCTGGA 123
Qy 121 agctctatttggcgtcggaccagagccatctatctgtgctgaggaagagcgtc 180
Db 124 AGCTCAATATTGGAGCAGCTGAACACGAGAGGTCATATCGGCATAGGAAGAGCGTT 183
Qy 181 gagggtctcagacacaaactgatttgggtgttctgtct 220
Db 184 GAAGGATCTCAGAACAAAGCTGATCTGTGTTTGTACT 223

RESULT 3

D22889
LOCUS D22889 403 bp mRNA EST 08-JUL-1999
DEFINITION R1C1720A Rice callus Oryza sativa cDNA clone C1720_1A, mRNA sequence.
ACCESSION D22889
VERSION D22889.1 GI:426816
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
AUTHORS Sasaki, T. and Minobe, Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.

FEATURES

source
location/Qualifiers

source

1..403
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="C1720_1A"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."
BASE COUNT 94 a 79 c 114 g 113 t 3 others
ORIGIN

Query Match 50.3%; Score 158.8; DB 36; Length 403;
Best Local Similarity 82.3%; Pred. No. 5.1e-39;
Matches 181; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 gataagtgccacactgagaagaagtaacctcccttgacatagaggttgatggtgt 60
Db 137 GAAAGGTGCGTCCCTTGAGAGAGAGTACCCATCCCTTGACATCGAGGTTGATGTTGT 196
Qy 61 ctaggctctcaacactagacgtgcccgcacatctgtcggggcccaattgcctgcctgga 120
Db 197 CTGGTCTTCCTCCACATGATGATGCGTGCATCTGCTGGGCTAACTGCTGCTGGA 256
Qy 121 agctctatttggcgtcggaccagagccatctatctgtcgtgaggaagagcgtc 180
Db 257 AGCTCAATATTGGAGCAGCTGAACACGAGAGGTCATATCGGCATAGGAAGAGCGTT 316
Qy 181 gagggtctcagacacaaactgatttgggtgttctgtct 220
Db 317 GAAGGATCTCAGAACAAAGCTGATCTGTGTTTGTACT 356

RESULT 4

C98335
LOCUS C98335 499 bp mRNA EST 19-OCT-1998
DEFINITION C98335 Rice callus Oryza sativa cDNA clone C1720_6Z, mRNA sequence.
ACCESSION C98335
VERSION C98335.1 GI:3761081
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
AUTHORS Sasaki, T. and Minobe, Y.
TITLE Rice cDNA from callus
JOURNAL Unpublished (1994)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.

FEATURES

source
location/Qualifiers
1..499

/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="C1720_6Z"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."
BASE COUNT 138 a 91 c 126 g 142 t 2 others
ORIGIN

[illegible]


```

POLYA=Yes.
FEATURES
  source
    Location/Qualifiers
      1..165
        /organism="Sorghum bicolor"
        /db_xref="taxon:4558"
        /clone_lib="Dark Grown 1 (DG1)"
        /note="Organ: 5-day-old dark-grown seedlings; Vector:
        Lambda zap; Site_1: XhoI; Site_2: EcoRI; The library was
        made from poly-A RNA in the cloning vector lambda ZAP II.
        Clones to be sequenced were prepared by mass excision."
BASE COUNT      41 a      46 g      45 t
ORIGIN
      1..165
        ctgggtcttcaaccatgagacgtggcgcatctgctggggccaattgcatgctgctgga 120
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        387 TTGGGGCCATCAACCACTTGATCTGCGAGCGGAAGCTGGAGCAAAATTCGATTTGTTCTGGG 446
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        121 agctctatatattgctgctgggacccagagccatcatctatctgctgaggaagagcgctc 180
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        447 AGTTCAGTGTGGTGGTCCCTGAAACAGCTCAAGTATCTCTGTTATTTAGGAGAGCGGTG 506
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        181 gagggctctcagacacaaaa 200
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        507 GAGCAAGGCCCGCGGTAACA 526
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Query Match      33.7%; Score 106.6; DB 35; Length 165;
Best Local Similarity 96.5%; Pred. No. 7.9e-23;
Matches 109; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gataaggtgcacactgagaagaagtacccttcccttgacatagaggttgatggtgt 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 53 GATAAGGTGCGCACATTGAGAAGAAGTACCCTTCCCTTGACATAGAGGTTGATGGTGT 112
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 ctaggctctcaaccatgagacgtggcgcatctgctggggccaattgcatcgt 113
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 113 TTAGGCCCTTCAACCATAGACGTGTCGCGCATCTGCTGGGGCCATTCATCGT 165
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
LOCUS      BE034126      749 bp      mRNA      EST      07-JUN-2000
DEFINITION MG05C09 MG Mesembryanthemum crystallinum cDNA 5' similar to
d-ribulose-5-phosphate 3-epimerase, mRNA sequence.
ACCESSION      BE034126
VERSION      BE034126.1 GI:8329120
KEYWORDS      EST.
SOURCE      common ice plant.
ORGANISM      Mesembryanthemum crystallinum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE      Bohnert, H.J., Borchert, C., Brooks, J., Eaton, M., Ferrea
AUTHORS      H., Kawasaki, S., McCollough, A., Michalowski, C.B., Palacio, C.,
Scara, G., Wheeler, M., and Zepeda, G.R.
Functional Genomics of Plant Stress Tolerance
Unpublished (2000)
Contact: Michalowski, C.B.
University of Arizona
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu
An open reading frame exists.
FEATURES
  source
    Location/Qualifiers
      1..749
        /organism="Mesembryanthemum crystallinum"
        /db_xref="taxon:3544"
        /clone_lib="MG"
        /tissue_type="roots"
        /dev_stage="5-6 weeks"
        /note="grown in hydroponics, stress 400 mM NaCl (in 0.5
        Hoagland's), 30 h stress"
BASE COUNT      207 a      194 g      213 t      2 others
ORIGIN
      1..749
        gataaggtgcacactgagaagaagtacccttcccttgacatagaggttgatggtgt 60
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        327 GATAAGGTTCGTCGTGAGGAGAGATCCATCTCTTGATATTGAGGTGATGGTGT 386
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Query Match      31.9%; Score 100.8; DB 33; Length 749;
Best Local Similarity 69.0%; Pred. No. 7.1e-21;
Matches 138; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 gataaggtgcacactgagaagaagtacccttcccttgacatagaggttgatggtgt 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 327 GATAAGGTTCGTCGTGAGGAGAGATCCATCTCTTGATATTGAGGTGATGGTGT 386
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 61 ctaggctcttcaaccatgagacgtggcgcatctgctggggccaattgcatgctgctgga 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 387 TTGGGGCCATCAACCACTTGATCTGCGAGCGGAAGCTGGAGCAAAATTCGATTTGTTCTGGG 446
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 agctctatatattgctgctgggacccagagccatcatctatctgctgaggaagagcgctc 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 447 AGTTCAGTGTGGTGGTCCCTGAAACAGCTCAAGTATCTCTGTTATTTAGGAGAGCGGTG 506
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 gagggctctcagacacaaaa 200
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 507 GAGCAAGGCCCGCGGTAACA 526
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
LOCUS      AI993279      483 bp      mRNA      EST      08-SEP-1999
DEFINITION 701495920 A. thaliana, Ohio State clone set Arabidopsis thaliana
cDNA clone 701495920, mRNA sequence.
ACCESSION      AI993279
VERSION      AI993279.1 GI:5840184
KEYWORDS      EST.
SOURCE      thale cress.
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE      1 (bases 1 to 483)
AUTHORS      Chen, J., Hillebrand, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
Gorgone, G., Burns, D., Griffin, J., Mouanoutou, M., Nguyen, D., Tan, R.,
Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.,
Policky, J., Suzuki, G., Argentine, C., Khan, S., Nobrrega, A., Murry, L.,
Turner, C., Krikorian, S., Elder, L. and Hanson, D.
Arabidopsis thaliana Gene Expression MicroArray
Unpublished (1999)
Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.
FEATURES
  source
    Location/Qualifiers
      1..483
        /organism="Arabidopsis thaliana"
        /db_xref="taxon:3702"
        /clone_lib="A. thaliana, Ohio State clone set"
        /note="cDNA library was made from selected clones from the
        Arabidopsis thaliana Ohio State clone set."
BASE COUNT      120 a      109 c      121 g      133 t
ORIGIN
      1..483
        gataaggtgcacactgagaagaagtacccttcccttgacatagaggttgatggtgt 60
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        31 GACAAAGTCCGGGCTTGGAGGATATACCAACACTTGATATTCAGGTGATGGCGGC 90
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        61 ctaggctcttcaaccatgagacgtggcgcatctgctggggccaattgcatgctgctgga 120
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        91 TTAGGCCCTTCAACGATCGTCACAGCGGCTGCAGCGCGCAACTGTATTGTTCTCTGGA 150
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        121 agctctatatattgctgctgggacccagagccatcatctatctgctgaggaagagcgctc 180
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        151 AGTTCAGTGTGGAGTCCCGAGCGGAGGATGTCATCTCCCTTCTGAGGAGCGGATGT 210
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
        181 gagggctctcagacacaaaaactgatt 206
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 211 GAGAAAGCCCAACCCACCACTTGAGT 236

RESULT 13

AV545174/c 544 bp mRNA EST 23-JUL-2000
 LOCUS AV545174 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
 DEFINITION cDNA clone R274e06F 3', mRNA sequence.
 ACCESSION AV545174
 VERSION AV545174.1 GI:8716588
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (bases 1 to 544)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)

JOURNAL

COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source
 1..544
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="R274e06F"
 /clone_lib="Arabidopsis thaliana roots Columbia"
 /tissue_type="roots"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 135 a 161 c 113 g 135 t
 ORIGIN

Query Match 30.18; Score 95; DB 19; Length 544;
 Best Local Similarity 69.94; Pred. No. 4.2e-19;
 Matches 128; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1 gataaggtgcgcacactgagaagaagtaaccttcccttgacatagaggttgatggtggt 60

Db 187 GACAAAGTTCAGGGCATTTGAGGAACAAGTACCCACACTTTGATATTGAGGTGGAGCGGCGC 128

QY 61 ctaggctctcaaccatagacgtggccgcatctgctggggcccaattgcatcgtcgctgga 120

Db 127 TTAGGCCCTTCCACAATCGATGACAGCTGCAGCTGGAGCACAACCTGATCGTCGACGGA 68

QY 121 agctctatttgcgctgcgacccagagccatcatctgtgctgaggaagagcgtc 180

Db 67 ACTTCAGTGTGTTGGAGTCCGAGCCCTGGGAGCGTCATATCCCTTTTGGCGGCTAGTGT 8

QY 181 gag 183

Db 7 GAG 5

RESULT 14

AI999119/c 619 bp mRNA EST 08-SEP-1999
 LOCUS AI999119 A. thaliana, Columbia Col-0, rosette-3 Arabidopsis
 DEFINITION thaliana cDNA clone 701554529, mRNA sequence.
 ACCESSION AI999119
 VERSION AI999119.1 GI:5846024
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

REFERENCE
AUTHORS

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 619)
 Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
 Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
 Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R.,
 Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.,
 Policky, J., Suzuki, G., Argentine, C., Shah, S., Nbrigga, A., Murry, L.,
 Turner, C., Krikorian, S., Elder, L. and Hanson, D.
 Arabidopsis thaliana Gene Expression MicroArray
 Unpublished (1999)

TITLE

JOURNAL

COMMENT

Contact: David Smoller, Ph.D.
 Genome Systems, Inc., a wholly owned subsidiary of Inceyte
 Pharmaceuticals, Inc.
 4633 World Parkway Circle, St. Louis, MO 63134, USA
 Tel: 877-577-2733
 Fax: 314-427-3324
 Email: service@genomesystems.com.

FEATURES

source

1..619
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia Col-0"
 /db_xref="taxon:3702"
 /clone="701554529"
 /clone_lib="A. thaliana, Columbia Col-0, rosette-3"
 /tissue_type="rosette"
 /dev_stage="4 - 7 weeks"
 /note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; cDNA
 library was derived from untreated rosette tissue from
 Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks.
 Plants were grown in 1:1:1 peat moss/vermiculite/perlite
 soil at 22 deg. C +/- 3 deg. C under constant light, and
 watered with fertilizer. cDNA synthesis was initiated
 using a NotI-oligo(dT) primer. Double-stranded cDNA was
 blunted, ligated to SalI adaptors, digested with NotI,
 size-selected, and cloned into the NotI and SalI sites of
 the pSPORT vector."

BASE COUNT 164 a 160 c 123 g 168 t 4 others
 ORIGIN

Query Match 29.9%; Score 94.6; DB 14; Length 619;
 Best Local Similarity 68.1%; Pred. No. 5.8e-19;
 Matches 130; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1 gataaggtgcgcacactgagaagaagtaaccttcccttgacatagaggttgatggtggt 60

Db 388 GACAAGGTGAGGCGCATTTGAGGAACAAGTACCCACACTTGATATTGAGGTGGAGCGGCGC 329

QY 61 ctaggctctcaaccatagacgtggccgcatctgctggggcccaattgcatcgtcgctgga 120

Db 328 TTAGGCCCTTCCACAATCGATGACAGCTGCAGCTGGAGCACAACCTGATCCTTCGACGGA 269

QY 121 agctctatttgcgctgcgacccagagccatcatctgtgctgaggaagagcgtc 180

Db 268 ACTTCAGTGTGTTGGAGTCCGAAAGCTGGGAGCGTCATATCCCTTTTGGCGGNTAGTGT 209

QY 181 gaggcctctca 191

Db 208 GAGAAAGCACA 198

RESULT 15

T75815 317 bp mRNA EST 09-JAN-1998
 LOCUS T75815 Lambda-PRL2 Arabidopsis thaliana cDNA clone 146H5T7, mRNA
 DEFINITION sequence.
 ACCESSION T75815
 VERSION T75815.1 GI:934873
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta: eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS 1 (bases 1 to 317)
Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Retzel
,E. and Somerville,C.

TITLE Genes galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cDNA clones
Plant Physiol. 106, 1241-1255 (1994)

JOURNAL
MEDLINE
COMMENT 95148729

On Apr 14, 1993 this sequence version replaced gi:692577.

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.

Lansing, MI

Tel: 517-353-0854

Fax: 517-353-9168

Email: 22313tcn@ibm.cl.msu.edu

Seq primer: T7 dye primer.

Location/Qualifiers

1..317

/organism="Arabidopsis thaliana"

/strain="var columbia"

/db_xref="taxon:3702"

/clone="146H5T7"

/clone_lib="Lambda-PRL2"

/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;

Lambda PRL2 is a cDNA library derived from equal

quantities of 4 pools of mRNA. The mRNA sources were 1) 7

day germinated etiolated seedlings; 2) tissue culture

grown roots; 3) staged plants half with 24 hour light

cycle, half on 16 hr light, 8 hour dark- rosettes; 4)

same plants as 3 but aerial tissue (stems, flowers and

siliques. The vector is BRL's lambda Zip-Lox. The cDNA

inserts were directionally cloned with Sal-Not arms using

oligo dt primed cDNA. "

BASE COUNT 74 a 68 C 85 g 79 t 11 others
ORIGIN

Query Match 28.5%; Score 90.2; DB 39; Length 317;
Best Local Similarity 64.6%; Pred. No. 1.2e-17;
Matches 148; Conservative 0; Mismatches 80; Indels 1; Gaps 1;

QY 1 gataaggtgcacactgagaagaagatccctcccttgacatagaggttaagtgtgt 60

Db 31 GACAAAGTCCGGGATTGAGGCAAAATACCAACACTTGATATTCAGGTGGAAGCGGC 90

QY 61 ctagtccctcaaccatagacgtgcccgcacatctgtggggcccaattgcctcgtgga 120

Db 91 TTAGGCCCTTCAACGATCGACACAGCGGCTGCACGCGGCCCACTGTATTGTTCTGGA 150

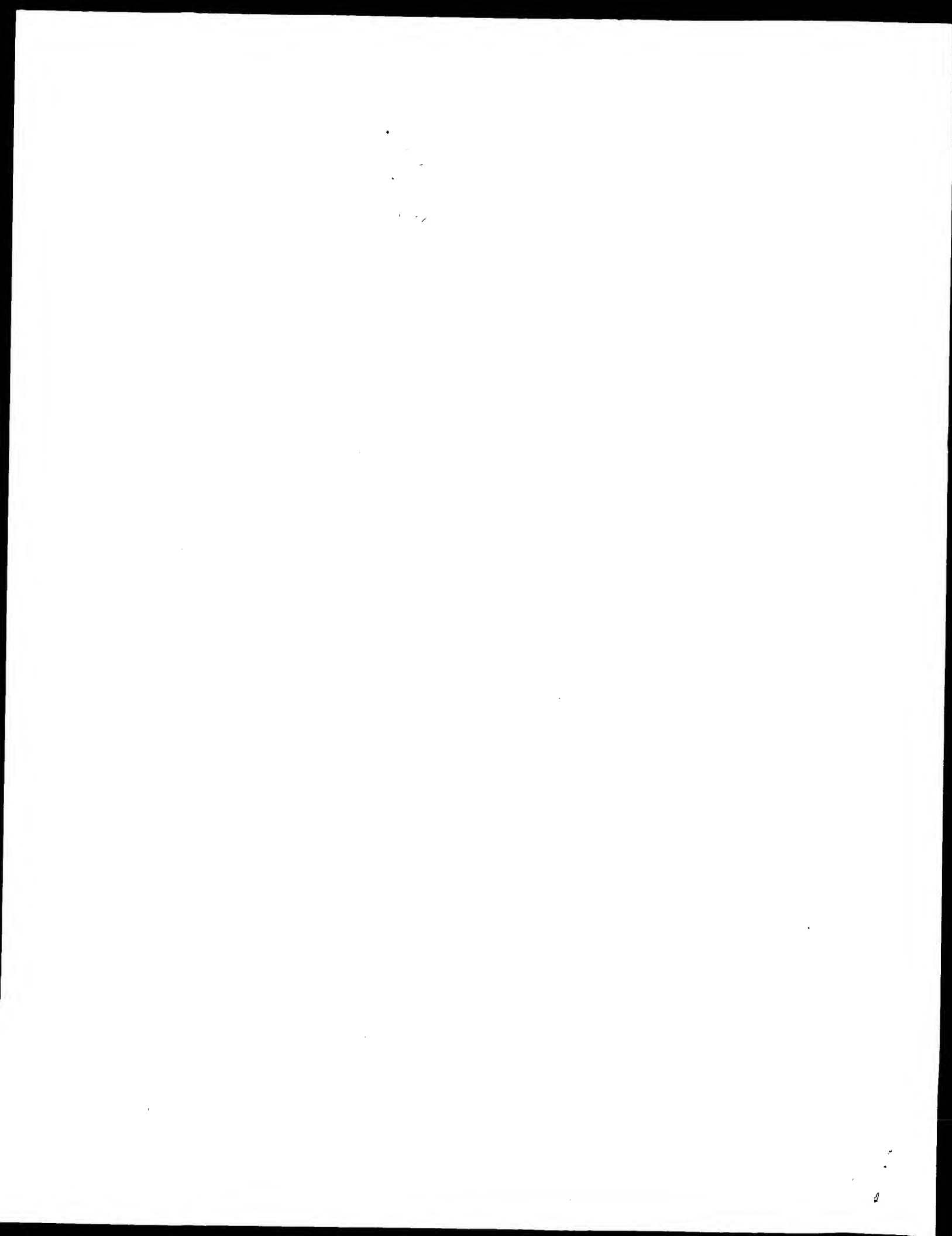
QY 121 agctctatatattggcgtgcgaccagagaccatcatat-ctgtgctgaggaagcgt 179

Db 151 AGTTTCAGTCTTTGGAGCTCCGGAGCGCAGGAGATGTCATCTCCCTTTCTNAGGACCACTGT 210

QY 180 cgagggtctcagaacaaactgatttgggtgtttctgtgtaagta 228

Db 211 TGAGAAAGCCCAACCCACCACCTTTAGTAGTACCACTTTTCTNAAATA 259

Search completed: November 4, 2000, 11:52:10
Job time: 18247 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:21 ; Search time 4352.3 Seconds
(without alignments)
426.175 Million cell updates/sec

Title: US-09-300-482-619

Perfect score: 300

Sequence: 1 gattattgacaacacgat.....ttcagaaatgtctgtgtgg 300

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_est3:*
- 4: gb_est4:*
- 5: gb_est5:*
- 6: gb_est6:*
- 7: gb_est7:*
- 8: gb_est8:*
- 9: gb_est9:*
- 10: gb_est10:*
- 11: gb_est11:*
- 12: gb_est12:*
- 13: gb_est13:*
- 14: gb_est14:*
- 15: gb_est15:*
- 16: gb_est16:*
- 17: gb_est17:*
- 18: gb_est18:*
- 19: gb_est19:*
- 20: gb_est20:*
- 21: gb_est21:*
- 22: gb_est22:*
- 23: gb_est23:*
- 24: gb_est24:*
- 25: gb_est25:*
- 26: gb_est26:*
- 27: gb_est27:*
- 28: gb_est28:*
- 29: gb_est29:*
- 30: gb_est30:*
- 31: gb_est31:*
- 32: gb_est32:*
- 33: gb_est33:*
- 34: gb_est34:*
- 35: gb_est35:*
- 36: gb_est36:*
- 37: gb_est37:*
- 38: gb_est38:*
- 39: gb_est39:*
- 40: gb_est40:*
- 41: em_estba:*
- 42: em_estfun:*
- 43: em_esthum1:*

44: em_esthum2:*

45: em_esthum3:*

46: em_esthum4:*

47: em_esthum5:*

48: em_esthum6:*

49: em_esthum7:*

50: em_esthum8:*

51: em_esthum9:*

52: em_esthum10:*

53: em_esthum11:*

54: em_esthum12:*

55: em_esthum13:*

56: em_esthum14:*

57: em_esthum15:*

58: em_esthum16:*

59: em_esthum17:*

60: em_esthum18:*

61: em_esthum19:*

62: em_esthum20:*

63: em_estin1:*

64: em_estin2:*

65: em_estin3:*

66: em_estin4:*

67: em_estom:*

68: em_estovi:*

69: em_estov2:*

70: em_estpl1:*

71: em_estpl2:*

72: em_estpl3:*

73: em_estpl4:*

74: em_estpl5:*

75: em_estro1:*

76: em_estro2:*

77: em_estro3:*

78: em_estro4:*

79: em_estro5:*

80: em_estro6:*

81: em_estro7:*

82: em_estro8:*

83: em_estro9:*

84: em_estro10:*

85: em_estro11:*

86: em_estro12:*

87: em_estro13:*

88: gb_gss1:*

89: gb_gss2:*

90: gb_gss3:*

91: gb_gss4:*

92: em_gss1:*

93: em_gss2:*

94: em_gss3:*

95: em_gss4:*

96: gb_gss5:*

97: gb_gss6:*

98: gb_gss7:*

99: gb_gss8:*

100: gb_gss9:*

101: em_gss5:*

102: em_gss6:*

103: em_gss7:*

104: em_gss8:*

105: em_gss9:*

106: em_gss10:*

107: em_gss11:*

108: gb_gss10:*

109: gb_gss11:*

110: em_gss12:*

111: gb_gss12:*

112: gb_gss13:*

113: gb_gss14:*

114: gb_gss15:*

115: gb_gss16:*

116: gb_gss17:*

117: gb_gss18.*
 118: gb_gss19.*
 119: em_gss13.*
 120: gb_gss20.*
 121: gb_gss21.*
 122: gb_gss22.*
 123: gb_gss23.*
 124: gb_gss24.*
 125: em_gss14.*
 126: em_gss15.*
 127: em_gss16.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	285.6	95.2	573	12	AI657311	AI657311 486093C03
2	225.2	75.1	423	23	AW564778	AW564778 LGI_301_C
3	223.8	74.6	330	36	D43256	D43256 D43256 Rice
4	210.4	70.1	550	34	BE205297	BE205297 EST397973
5	208.8	69.6	419	19	AW035534	AW035534 EST281272
6	191.4	63.8	670	23	AW584861	AW584861 N211008e
7	182.4	60.8	538	25	AW933758	AW933758 EST359601
8	172	57.3	537	22	AW399786	AW399786 EST310286
9	135.8	45.3	597	35	BE443964	BE443964 WHE1123_E
10	133	44.3	217	14	AL366975	AL366975 MCB1123_E
11	106.6	35.5	501	21	AW254881	AW254881 ML1306 pe
12	104	34.7	621	21	AW255839	AW255839 ML923 pep
13	94.6	31.5	564	11	AI482995	AI482995 EST242318
14	89	29.7	439	35	BE400087	BE400087 AWB010.F1
15	83	27.9	588	18	AW522495	AW522495 AV522495
16	71.6	23.9	681	21	AW256424	AW256424 EST304491
17	71	23.7	452	113	AQ953823	AQ953823 nbe0053A
18	61	20.3	404	37	L33519	L33519 BNAESTF092
19	58.6	19.5	326	19	AV551048	AV551048 AV551048
20	42.8	14.3	465	114	AZ046382	AZ046382 nbe00085N
21	39.8	13.3	533	23	AW569848	AW569848 s182802.Y
22	37.8	12.6	506	90	AQ240320	AQ240320 CIT-HSP-2
23	34	11.3	240	15	AU074940	AU074940 AU074940
24	34	11.3	704	15	AU061833	AU061833 AU061833
25	33.6	11.2	254	22	AW484099	AW484099 59215 MAR
26	33.6	11.2	1045	124	CNS05JIG	AL340225 Tetraodon
27	33.4	11.1	266	36	BE529610	BE529610 M75E07STM
28	33.2	11.1	416	37	F15380	F15380 ATT55628 G1
29	33.2	11.1	546	9	A1260448	A1260448 LP04319.5
30	32.6	10.9	946	121	CNS006RN	AL065838 Drosophila
31	32.4	10.8	369	22	AW430628	AW430628 70368 MAR
32	32.2	10.7	458	117	AZ281677	AZ281677 RPCI-23-1
33	32.2	10.7	611	114	AZ064575	AZ064575 RPCI-23-4
34	32	10.7	191	29	BB273748	BB273748 BB273748
35	32	10.7	851	121	CNS00AEL	AL055177 Drosophila
36	31.6	10.5	334	9	A1217424	A1217424 qb22h09.x
37	31.6	10.5	925	121	CNS00LEK	AL068320 Drosophila
38	31.4	10.5	369	3	AA327905	AA327905 EST31307
39	31.4	10.5	431	8	A1034194	A1034194 ox12h07.x
40	31.4	10.5	451	10	A1452153	A1452153 mp80h08.x
41	31.4	10.5	468	20	AW221197	AW221197 EST297666
42	31.2	10.4	372	33	BE109286	BE109286 UI-R-BJ1-
43	31.2	10.4	399	22	AW531748	AW531748 UI-R-C4-a
44	31.2	10.4	457	1	AA004451	AA004451 zh89c05.s
45	31.2	10.4	492	32	BE020509	BE020509 sm44f04.Y

ALIGNMENTS

RESULT 1
 AI657311

LOCUS AI657311 573 bp mRNA EST 05-MAY-1999
 DEFINITION 486093C03.y1 486 - leaf primordia cDNA library from Hake lab Zea mays cDNA, mRNA sequence.
 ACCESSION AI657311
 VERSION AI657311.1 GI:4753401
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 REFERENCE Walbot.V.
 AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford University
 TITLE Unpublished (1999)
 JOURNAL Contact: Walbot V
 COMMENT Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 486093 row: C column: 03.
 FEATURES
 source
 1..573
 /location="Qualifiers
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="486 - leaf primordia cDNA library from Hake lab"
 /issue_type="leaf primordia"
 /dev_stage="p7-p11 leaf"
 /lab_host="E. coli XL1-Blue MFR"
 /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA library."
 BASE COUNT 161 a 115 c 133 g 164 t
 ORIGIN
 Query Match 95.2%; Score 285.6; DB 12; Length 573;
 Best Local Similarity 97.0%; Pred. No. 4.1e-79;
 Matches 291; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 gatttattgcaacacgcagctctgtggttgatgatcatcaaatgtctcaactaggacctg 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 GATTTATTGACACACCGATCTCTGCTGGGATTGATCATCAAAATTGCTCAACTAGGACCTG 120
 Qy 61 aactggcaactactcttctgttaattgttcattcttaagcgaggagggcacactgaaacccgca 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 AACTGGCACTACTCTTGTAAATTGTCATTTCTTAAGAGCGGAGGACACCTGAAACCCGCA 180
 Qy 121 atggtctactagaagtacagaagcccttcagagatcggggctgcaattctctgaaacagg 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 ATGGGGCTACTGGGAAGTACAGAAAGCCCTTCAGAGATGCGAGTCTGCAATTCGAAACAGG 240
 Qy 181 gtgttgcattactcaagaataattctctgttgataacactgctagataagagggatgggt 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 GTGTTGCAATCACTCAAGAAATTTCTGTGTTGATAACACTGCCAGAAATAGAGGGATGTT 300
 Qy 241 tagctcgggttctctatgttttgattgggttggttggttaggaacttcagaaatgtctgctgg 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 TAGTCGGTTTCTCTATGTTTACTGGGTTGGTGGTAGGACTTCAGAAATGTCAGCTGTTG 360
 RESULT 2
 AW564778 423 bp mRNA EST 19-JUL-2000
 LOCUS LGI_301_C06.bl_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA
 DEFINITION sequence.
 ACCESSION AW564778
 VERSION AW564778.1 GI:7218656
 KEYWORDS EST.
 SOURCE sorghum.

ORGANISM

Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.

REFERENCE

1 (bases 1 to 423)

AUTHORS

Cordonnier-Pratt M.-M., Gingle, A., Marsala, C. and Pratt, L.H.

TITLE

An EST database from Sorghum: light-grown seedlings

JOURNAL

Unpublished (2000)

COMMENT

Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: JEN REV

High quality sequence stop: 420

POLYA-No.

FEATURES

Location/Qualifiers

source

1..423

/organism="Sorghum bicolor"

/db_xref="taxon:4598"

/clone_lib="Light Grown 1 (LG1)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI"

; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 114 a 110 c 93 g 106 t

ORIGIN

Query Match 75.18; Score 225.2; DB 23; Length 423;

Best Local Similarity 98.78; Pred. No. 3.4e-60;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y 1 gatttattgacacaccgacatcctgctggattgacatcaaatgctcaactaggacctg 60

Db 194 GATTATTGACACACCGATCTCTGCTGGTATTGATCATCAATTCCTCACTAGGACCTG 253

Y 61 aactggcaactactcttgtaattgtcattttcttaagagcggaggcacacctgaaacccgca 120

Db 254 AACTGGCAACTACTCTTGTAAATTCATTTCTAAGAGCGGAGGCACACCTGAAACCGCA 313

Y 121 atgtctactagaagtacagaagccttcagagatgcggggctgcaattctcgaacagg 180

Db 314 ATGCTCTACTGGAAGTACAGAAAGCCCTTCAGAGATCGCGGGCTGCAATTCGAAACAGG 373

Y 181 gtgttgaattactcaagaaattctctgttgataacactgctagaata 230

Db 374 GTGTGCAATTACTCAAGAAATTCCTCTGTGTGATACACTGCAGAAATA 423

RESULT 3

D43256

LOCUS

D43256 330 bp mRNA EST 04-MAY-1998

DEFINITION D43256 Rice callus cDNA (H.Uchimiya) Oryza sativa CDNA Clone SSI31, mRNA sequence.

ACCESSION D43256

VERSION D43256.1 GI:3107516

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.

1 (bases 1 to 330)

Uchimiya, H.

On nucleotide sequence of Oryza sativa

Unpublished (1994)

Contact: Hirofumi Uchimiya

Institute of Mol. & Cell. Bioscience, Department of Cellular

Function

The University of Tokyo

1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan

Tel: 03-3812-2111(ex.7844)

Fax: 03-3812-2910

Email: huchimiya@tansei.cc.u-tokyo.ac.jp

PROJECT= Uchimiya.

Location/Qualifiers

1..330

/organism="Oryza sativa"

/db_xref="taxon:4530"

/clone="SS131"

/clone_lib="Rice callus CDNA (H.Uchimiya)"

/tissue_type="callus"

BASE COUNT 93 a 64 c 84 g 89 t

ORIGIN

Query Match 74.68; Score 223.8; DB 36; Length 330;

Best Local Similarity 88.48; Pred. No. 8.8e-60;

Matches 243; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Y 26 tggattgatcatcaaatgctcaactagacacctggaactggaactactctgtgaattgt 85

Db 1 TGGAAATCGGACCATCAAAATTCCTCACTAGGACGAGCTTGCACTTCTTGAATTGT 60

Y 86 cattcttaagagcggaggcacacctgaaacccgcaatgctctactagaagtaacagaagc 145

Db 61 TATTCTAAGAGTGGAGGACACCTGAACCCGGAATGGTCTGCTTGAAGTCAAAAGC 120

Y 146 ctccagagatgcggggctgcaattctcgaaacagggtgttgcattactcaagaaattc 205

Db 121 CTTCCGAGATGCTGGGCTGGATTCTCGAAACAGGAGTTCGAGTTACTCAAGAAATTC 180

Y 206 tctgttggaataacactgctagagaggtggttagctcggttctcctatgttgattg 265

Db 181 TCTGTTAGACACACTGCCAGAAATAGAGGATGGTTAGCTAGATTTCCCATGTTGACTG 240

Y 266 ggttggtgtaggaacttcagaaatgtctgtgtg 300

Db 241 GGTGTGGAAGAATCATCAGAAATGTCACCTGCTCG 275

RESULT 4

BEZ05297

LOCUS

BEZ05297 550 bp mRNA EST 27-JUN-2000

DEFINITION EST397973 KV0 Medicago truncatula cDNA clone PKV0-21015, mRNA sequence.

ACCESSION BEZ05297

VERSION BEZ05297.1 GI:8748593

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.

1 (bases 1 to 550)

Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, J.S., Peng, H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.

ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula

Unpublished (2000)

Contact: Carroll P. Vance

Department of Agronomy and Plant Genetics

University of Minnesota

411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA

Tel: 612-625-7219

Fax: 651-649-5038

Email: vance004@maroon.tc.umn.edu

University of Minnesota name: T265069e

TIGR sequence name: MTGBK92TK

More information is available at.

D43256 EST Db

Sat Nov 4 18:11:44 2000

http://chrysie.tamu.edu/medicago
Seq primer: Skmod (CTA gAA CTA gTg gAT CC).

FEATURES

source
1. .550
/location/Qualifiers
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="PKV0-21015"
/clone_lib="KV0"
/tissue_type="Seedling roots"
/dev_stage="Immediately prior to inoculation with
Sinorhizobium meliloti (0 hour)"
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The
cDNA was directionally ligated into the Unizap XR vector
from Stratagene and packaged using Gigapack III Gold
packaging extracts. Plasmids containing cDNA inserts
were excised from the recombinant lambda-Zap phage using
Ex-assist helper phage and propagated in SOLR cells."
BASE COUNT 143 a 115 c 139 g 153 t
ORIGIN

Query Match 70.1%; Score 210.4; DB 34; Length 550;
Best Local Similarity 81.3%; Pred. No. 1.6e-55;
Matches 244; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 1 gatttattgacaacaccgacatcctgctggcattgacatcaaatgctcaactaggacctg 60
Db 126 GATTATTGACAACACAGATCCTGCTGGCATTGATCATCATGATTCGACAACTCGCCCTG 185
QY 61 aactggcaactactcttgaattgcttcttaagagcgaggacacacctgaaccgcga 120
Db 186 AGCTAGCTTCGACACTTGAATTTGATTTCAAGAGTGGAGGACCCCTCGAGACTAGAA 245
QY 121 atggtctactagaagtagacagaagccttcagagatgcgggctgcaattctcgaacagg 180
Db 246 ATGGCTTGTAGAACTACAAAGGCTTTCGTGAAGCAGGATTGGATTCCCAAAACAGG 305
QY 181 gtgttgcaattactcaagaaattctctgttggaataacactgctagaaatagaggatgg 240
Db 306 GTGTGCTATAACACAAAGAACTCTTGTGGATAAACACTGCCAGAAATTGAGGGATGGT 365
QY 241 tagctcggttctactgttgaattggttggtgtagaacttcaagaatgctctgctg 300
Db 366 TGGCTAGATTCCCATGTTGACTGGGTGGAGGTAGAACGTCAGAGATGCTGCGATCG 425

RESULT 5
AW035534 419 bp mRNA EST 15-SEP-1999
LOCUS
DEFINITION EST281272 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEC39G17 similar to glucose-6-phosphate isomerase, putative, mRNA
sequence.
ACCESSION AW035534
VERSION AW035534.1 GI:5894290
KEYWORDS EST.
SOURCE Tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 419)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Opton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU

FEATURES

source
1. .419
/location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC39G17"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
BASE COUNT 109 a 92 c 102 g 116 t
ORIGIN

Query Match 69.6%; Score 208.8; DB 19; Length 419;
Best Local Similarity 81.0%; Pred. No. 4.8e-55;
Matches 243; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 1 gatttattgacaacaccgacatcctgctggcattgacatcaaatgctcaactaggacctg 60
Db 112 GATTATTGACAACATACAGATCCAGCAGGCAATGATCATCAATTCGCAACACTGGCCCTG 171
QY 61 aactggcaactactcttgaattgcttcttaagagcgaggacacacctgaaccgcga 120
Db 172 AGCTGGCTTCAACACTGTTATTGTTATTTCAAGAGTGGAGGCACTCCAGAAACCCGAA 231
QY 121 atggtctactagaagtagacagaagccttcagagatgcgggctgcaattctcgaacagg 180
Db 232 ATGGCTTGTGTAAGTTTCAGAGGCTTCGCTGAAGCTGGCCTGATTTATTCGCAAAACAGG 291
QY 181 gtgttgcaattactcaagaaattctctgttggaataacactgctagaaatagaggatgg 240
Db 292 GTGTGCTATACACAAAGAACTCATTTGTTGACACACTGCTAGAAATTGAGGGTGGT 351
QY 241 tagctcggttctactgttgaattggttggtgtaggacctcagaatgctctgctg 300
Db 352 TGGCAGATTCCCATGTTGATTGGTGGCGGAGAACCTCAGAAATGCTGCGAGTTG 411

RESULT 6
AW584861 670 bp mRNA EST 28-APR-2000
LOCUS
DEFINITION N211008e MHAM Medicago truncatula/Glomsu versiforme mixed EST
library cDNA clone MHAM-7H23, mRNA sequence.
ACCESSION AW584861
VERSION AW584861.1 GI:7261915
KEYWORDS EST.
SOURCE Medicago truncatula/Glomsu versiforme mixed EST library.
ORGANISM Medicago truncatula/Glomsu versiforme mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 670)
AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after colonization with
Glomsu versiforme
JOURNAL Unpublished (2000)
COMMENT Contact: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org

Other name: MHAM-7c-D12; Date: 3/14/00; Updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at 'http://chrystie.tamu.edu/medicago'.

Seq primer: T3.

FEATURES

Location/Qualifiers
1..670
/organism="Medicago truncatula/Glomus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:110992"
/clone="MHAM-7H23"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in SOLR cells."

BASE COUNT 165 a 126 c 187 g 192 t
ORIGIN

Query Match 63.8%; Score 191.4; DB 23; Length 670;
Best Local Similarity 81.3%; Pred. No. 1.6e-49;
Matches 222; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 1 gattattgacacaccgacatctgtggtgattgatcatcaaatgtgctcaactagacatg 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 398 GATTTATTGACACACAGATCTGCTGGCATTTGATCATCATGATGACACAACTCGGCCCTG 457

QY 61 aactggcaactactgttaattgtcattcttaagagcggaggacacactgaacccgca 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 458 ACTAGCTTCGACACTTGTATTTGATTTCAAGAGTGGAGGCCACCCCTGAGACTAGAA 517

QY 121 atggtctactagaagtcagaaagccttcagagatgcgggctgcaattctcgaaacagg 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 518 ATGGCTTGTAGAGTACAAAGGCTTTCGTGAGGAGGATTTGGATTTCCCAAAACAGG 577

QY 181 gcttgcaattactcaagaaattctgttggtataacactgctagaaatagaggatggt 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 578 GTGTTGCTATACACAAGAAACACTTTTGTGGATAACACTGCCAGAAATTCAGGGATGGT 637

QY 241 tagctcggttctctatgtttgattggttggtg 273
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 638 TGGCTAGATTTCCCATGTTTACTGGTGGGGAG 670

RESULT 7
AW933758 538 bp mRNA EST 30-MAY-2000
LOCUS EST359601 tomato fruit mature green, TAMU Lycopersicon esculentum
DEFINITION cDNA clone cLEF56016 5', mRNA sequence.
ACCESSION AW933758
VERSION AW933758.1 GI:8109159
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.

REFERENCE 1 (bases 1 to 538)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T.S., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue

JOURNAL COMMENT

Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.

FEATURES

Location/Qualifiers
1..538
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF56016"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: XhoI; cLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp"

BASE COUNT 138 a 108 c 145 g 147 t
ORIGIN

Query Match 60.8%; Score 182.4; DB 25; Length 538;
Best Local Similarity 80.7%; Pred. NO. 1e-46;
Matches 213; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 37 atcaaatgtctcaactagagacctgaaactgcaactactcttgaattgtcattctaaaga 96
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATCAAAATTGCACAACTTGGCCCTGAGCTGCTTCACACTGTATTGTATTTCAAAGA 60

QY 97 gcggaggcacacactgaaacccgcaatggtctactagaagtcagaaagccttcagagatg 156
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GTGAGGACACTCCAGAAACCCGAAATGGCTTGTAGAAAGTTCAGAAAGCCCTTCGCTGAAG 120

QY 157 cggggctgcaattctcgaaacagggtgtgtcgaattactcaagaaattctctgttgata 216
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CTGCCCTGATATTTCGCAAAACAGGGTGTGCCATTACACAAGAAACTATTCTCTTGACA 180

QY 217 acactgtagaataagaggatggttagctgcttctctatgattgttgggttggtgata 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 ACACCTGCTAGAAATGAGGGTGTGCTGCCAGATTCCTCATGTTGATTGGTGGGCGGAA 240

QY 277 ggaactcagaataatgctgtgtgtg 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GAACCTCAGAAATGCTGCAGTTG 264

RESULT 8

AW999786 537 bp mRNA EST 07-FEB-2000
LOCUS EST310286 L. pennellii trichome, Cornell University Lycopersicon
DEFINITION pennellii cDNA clone cLTP10F14 5', mRNA sequence.
ACCESSION AW999786
VERSION AW999786.1 GI:6918256
KEYWORDS EST.
SOURCE Lycopersicon pennellii.
ORGANISM Lycopersicon pennellii
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 537)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E., Liang,F., Hansen,T.S., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from wild tomato (Lycopersicon pennellii)

TITLE

Niebel, A., Carreau, V., Chatagnier, O., Kahn, D., Gianinazzi-Pearson V. and Gamas, P.
Medicago truncatula ESTs from nitrogen-starved roots
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
MT-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Mtruncatula.html).
Location/Qualifiers
1. .217
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MCBAILD05"
/clone_lib="MtBA"
/tissue_type="root tips"
/dev_stage="harvested after 3 days of N-starvation"
/note="vector: pBluescript pSK; Site.1: EcoRI; Site.2:
XhoI; Plants were grown in an aeroponic chamber for 14
days on nitrogen-rich medium followed by 3 days on N-free
medium. RNA was extracted from root tips (1-3 cm). cDNA
was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zapR vector from
Stratagene and packaged using GigaPack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using ExSaxit helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."
BASE COUNT 58 a 37 c 60 g 61 t 1 others
ORIGIN

Query Match 44.3%; Score 133; DB 14; Length 217;
Best Local Similarity 78.4%; Pred. No. 2.5e-31;
Matches 171; Conservative 0; Mismatches 46; Indels 1; Gaps 1;
QY 58 ctgaactggaactactcttgaattgtctattcttaagagcggagggacacactgaaccc 117
Db 1 CTGAGCTAGCTTCGACACTTGTGATGTGATTTCAAAGAGTGGAGGCCCTTACACTA 60
QY 118 gcaatggtctactagaagcagcctcagagatcgcggtcgcaattctcgaac 177
Db 61 GAATGGCTTGTAGAGT-CAAAAGGCTTCGTGAAGCAGGATTTGCCAANAAC 119
QY 178 aggtgttgcaactactcaagaaaattctctgttggtataacactgtctagaatagggat 237
Db 120 AGGGTGTCTATTACACAAGAAACTCTCTTGTGTAACACTGCCAGATTGAGGGAT 179
QY 238 ggttagctcggttctctatgttattgttggttggtggt 275
Db 180 GGTGGCTAGATTTCCTATGTTTGAAGTGGTGGGAGGT 217

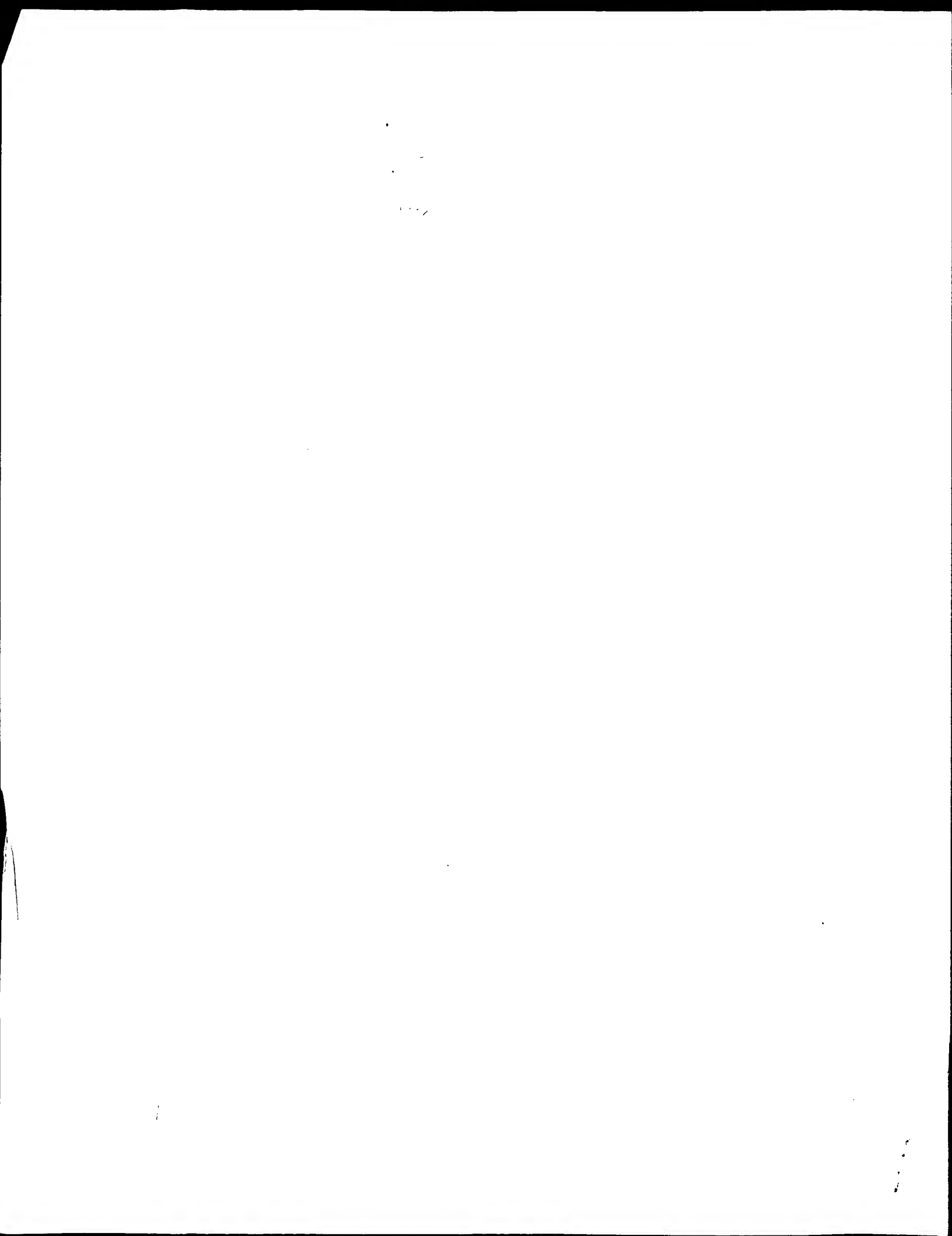
RESULT 11
AW254881 501 bp mRNA EST 15-MAR-2000
LOCUS
DEFINITION
ML1306 peppermint glandular trichome Mentha x piperita cDNA, mRNA
sequence.
ACCESSION
AW254881 1 GI:7244261
KEYWORDS
EST.
SOURCE
peppermint.
ORGANISM
Mentha x piperita
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Lamiales; Lamiaceae; Mentha.
REFERENCE
1 (bases 1 to 501)

Lange, B.M., Wildung, M.R., Stauber, E.J., Sanchez, C., Pouchnik, D. and
Croteau, R.
Probing essential oil biosynthesis and secretion by functional
evaluation of expressed sequence tags from mint glandular trichomes
Proc. Natl. Acad. Sci. U.S.A. 97 (6), 2934-2939 (2000)
Contact: Lange, B.M.
Institute of Biological Chemistry/Washington State University
Pullman, WA
Email: lange-m@mail.wsu.edu.
Location/Qualifiers
1. .501
/organism="Mentha x piperita"
/cultivar="Black Mitcham"
/db_xref="taxon:34256"
/clone_lib="peppermint glandular trichome"
/tissue_type="pellate glandular trichomes"
/cell_type="secretory"
/note="Vector: lambda ZAPII"
BASE COUNT 145 a 105 c 129 g 122 t
ORIGIN

Query Match 35.5%; Score 106.6; DB 21; Length 501;
Best Local Similarity 76.9%; Pred. No. 6.1e-23;
Matches 130; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 gatttattgacacacgcgctctgttggttgatcatcaaatgtctcaactaggacctg 60
Db 333 GATTTCATTGACATACAGATCCAGCGGGATTGATCATGATAGCACCACTTGGCAGTG 392
QY 61 aactgcaactactcttgaattgtctattctaaagcggagcagcactgaacccgca 120
Db 393 AGCTAGATCTACACTCGCTGATGATGTTTCAAGAGTGGAGTACGCCGGAACAGAA 452
QY 121 atgtctactagaagtacagaaagccttcagagatcgcggtggtcgcaatt 169
Db 453 ACGGTCTACTTGAAGTTTCAGAAAGCCTTCGGAAGAGCTGCTGTGGATT 501

RESULT 12
AW255839 621 bp mRNA EST 15-MAR-2000
LOCUS
DEFINITION
ML923 peppermint glandular trichome Mentha x piperita cDNA, mRNA
sequence.
ACCESSION
AW255839 1 GI:7245091
KEYWORDS
EST.
SOURCE
peppermint.
ORGANISM
Mentha x piperita
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Lamiales; Lamiaceae; Mentha.
REFERENCE
1 (bases 1 to 621)
Lange, B.M., Wildung, M.R., Stauber, E.J., Sanchez, C., Pouchnik, D. and
Croteau, R.
Probing essential oil biosynthesis and secretion by functional
evaluation of expressed sequence tags from mint glandular trichomes
Proc. Natl. Acad. Sci. U.S.A. 97 (6), 2934-2939 (2000)
Contact: Lange, B.M.
Institute of Biological Chemistry/Washington State University
Pullman, WA
Email: lange-m@mail.wsu.edu.
Location/Qualifiers
1. .621
/organism="Mentha x piperita"
/cultivar="Black Mitcham"
/db_xref="taxon:34256"
/clone_lib="peppermint glandular trichome"
/tissue_type="pellate glandular trichomes"
/cell_type="secretory"
/note="Vector: lambda ZAPII"
BASE COUNT 181 a 138 c 156 g 145 t
ORIGIN

Search completed: November 4, 2000, 11:52:25
Job time: 18262 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:34:08 ; Search time 189.35 Seconds
(without alignments)
307.515 Million cell updates/sec

Title: US-09-300-482-356

Perfect score: 385

Sequence: 1 caaccggcaccagcctgac.....acggtctggaggttttttt 385

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents_NA.*
2: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
8: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	170.4	44.3	2629	4	US-09-012-030-1
2	170.4	44.3	2629	4	US-08-590-454-1
3	76.8	19.9	1995	6	PCT-US96-05320A-1025
4	51.6	13.4	1872	4	US-08-743-6378-17
5	51.6	13.4	1872	5	US-08-526-8408-17
6	41.2	10.7	4195	2	US-08-340-011-1
7	41.2	10.7	4195	5	US-08-901-710-1
8	41.2	10.7	4425	1	US-08-222-615-31
9	41.2	10.7	4425	6	PCT-US95-04228-31
10	41.2	10.7	4795	2	US-08-340-011-3
11	41.2	10.7	4795	5	US-08-901-710-3
12	41.2	10.7	9108	6	PCT-US95-04228-45
13	41	10.6	1693	5	US-09-320-878-23
14	36.2	9.4	1931	4	US-09-130-114-2
15	35.6	9.2	1136	2	US-08-106-981-5
16	35.6	9.2	1431	1	US-08-254-357-1
17	35.6	9.2	1632	1	US-07-959-941-1
18	35.6	9.2	1632	1	US-08-259-924-1
19	35.6	9.2	4983	1	US-08-472-358-1
20	35.6	9.2	4983	6	PCT-US92-05786A-1
21	35.4	9.2	68750	5	US-09-335-409-1
22	34.8	9.0	2238	2	US-08-742-011-1
23	34.6	9.0	1157	1	US-08-095-726-1
24	34.6	9.0	1157	1	US-08-095-726-3
25	34.6	9.0	1157	1	US-08-096-043-1
26	34.6	9.0	1157	1	US-08-096-043-3

27	34.6	9.0	1157	1	US-08-093-577-1	Sequence 1, Appli
28	34.6	9.0	1157	1	US-08-093-577-3	Sequence 3, Appli
29	34.6	9.0	1157	1	US-08-096-623A-1	Sequence 1, Appli
30	34.6	9.0	1157	1	US-08-096-623A-3	Sequence 3, Appli
31	34.6	9.0	50341	2	US-08-247-901C-1	Sequence 1, Appli
32	34.6	9.0	50341	4	US-09-075-904-1	Sequence 1, Appli
33	34.4	8.9	1513	1	US-08-314-309A-2	Sequence 2, Appli
34	34.4	8.9	1546	1	US-08-314-309A-3	Sequence 3, Appli
35	34.4	8.9	1621	2	US-08-722-001-13	Sequence 13, Appli
36	34.4	8.9	1776	2	US-08-722-001-29	Sequence 29, Appli
37	34.4	8.9	2140	1	US-08-334-698-1	Sequence 1, Appli
38	34.4	8.9	2140	1	US-08-228-932-1	Sequence 1, Appli
39	34.4	8.9	2140	2	US-08-468-939-1	Sequence 1, Appli
40	34.4	8.9	2140	3	US-08-406-855A-1	Sequence 1, Appli
41	34.4	8.9	2140	4	US-08-722-190-1	Sequence 1, Appli
42	34.4	8.9	2140	5	US-08-244-354-1	Sequence 1, Appli
43	34.4	8.9	2140	5	US-09-206-899-1	Sequence 1, Appli
44	34.4	8.9	2140	6	PCT-US95-04203-1	Sequence 1, Appli
45	34.4	8.9	3172	1	US-08-314-309A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-012-030-1
; Sequence 1, Application US/09012030
; Patent No. 5912169
; GENERAL INFORMATION:
; APPLICANT: SCHMIDT, Ralf-Michael, STITT, Marc, SONNEWALD,
; APPLICANT: Uwe
; TITLE OF INVENTION: Transketolase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage.
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,030
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/590,454
; FILING DATE: 22-JAN-1996
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana
; FEATURE:
; NAME/KEY: CDA
; LOCATION: 60..2289
US-09-012-030-1

Query Match 44.3%; Score 170.4; DB 4; Length 2629;
Best Local Similarity 67.4%; Pred. No. 3.6e-37;
Matches 240; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 4 ccggcaccagcctgacatcattgggttgccaccgctccgagctgcgggca 63

; TOPOLOGY: linear
PCT-US96-05320A-1025[illegible]

RESULT

US-08-743-637B-17
: Sequence 17, Application US/08743637B
: Patent No. 5994066
: GENERAL INFORMATION:
: APPLICANT: BERGERON, Michel G.
: APPLICANT: PICARD, Francois J.
: APPLICANT: OUELLETTE, Marc
: APPLICANT: ROY, Paul H.
: TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
: TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
: TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
: NUMBER OF SEQUENCES: 273
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: QUARLES & BRADY
: STREET: 411 EAST WISCONSIN AVENUE
: CITY: MILWAUKEE
: STATE: WISCONSIN
: COUNTRY: USA
: ZIP: 53202-4497
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/743,637B
: FILING DATE: 04-NOV-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/526,840
: FILING DATE: 11-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: BAKER, Jean C.
: REGISTRATION NUMBER: 35,433
: REFERENCE/DOCKET NUMBER: 850586.90012
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (414) 277-5000
: TELEFAX: (414) 277-5591

```

: INFORMATION FOR SEQ ID NO: 17:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1872 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: ORIGINAL SOURCE:
:   ORGANISM: pseudomonas aeruginosa
US-08-743-637B-17

```

Query Match	13.4%	Score 51.6;	DB 4;	Length 1872;
Best Local Similarity	47.3%;	Pred. No. 2.9e+05;		
Matches 156;	Conservative 0;	Mismatches 174;	Indels 0;	Gaps 0;
QY 4	ccggcacaagcctgacatcaattgggttgggcaaccggtccgagctcgagatcgaggagccgga	63		
Db 716	CCGCTTCGGGCGGTACTGCTGTGTGATCGGCCGCGACGGCGGATCGACATGATCGACC	775		
QY 64	atgcgcgcgacgagctgagaaaggaggggaagcgggccgcgtgctctgcttctctct	123		
Db 776	TGTGGGCCAAGGACGCCACCAAGTCCGCCGAGATCAAGATCGGGATCGAGCGCGCTCGG	835		
QY 124	gggaactctttgatgacgtcgagtagtacaaggagagcgtctcctcctgccagactca	183		
Db 836	TGGAAGGCTCCAAAGGGCTACGAGGACCGCTACACCATCGCGCGCGCTACTGGC	895		
QY 184	cagcagagatcagatcagagcgcggtctcactctcgctggcagaagaatcagtcggagcc	243		
Db 896	CGCCGCAGTTCGGGATCATGGACGGCGAGACCCCTTGGAAACGAAGACAGATGCTTCCACCC	955		
QY 244	aaggcaaggccattggcatcgacaagtctgcgcgagtgctcctgcgggagcagctaca	303		
Db 956	CGGSCATGACCGTAGACACCCAGACCTTACCACCCGGGAACCGCGCGTGGCGGATCATCG	1015		
QY 304	aggaagtcagcgcacccgtggagacatca	333		
Db 1016	CTCTCCACGAGCACCCCGAGTTCATCTGCA	1045		

RESULT

RESOLVED
US-08-526-840B-17
Sequence 17, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE
TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994

```
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas aeruginosa
; US-08-526-840B-17

Query Match          13.4%; Score 51.6; DB 5; Length 1872;
Best Local Similarity 47.3%; Pred. No. 2.9e-05;
Matches 156; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

Qy 4 ccggcaccagcctgacatcatgttggtggccaccggctccgagctggagatcgcgggca 63
Db 716 CCGCTTCGGCGCGCTACTGCTGTGTGATCGCGCGACGCGCGGATCGACATGATCGACC 775

Qy 64 atgcgccgacagctgaggaagagggggaagcgtcccgctgctgtctgtctctct 123
Db 776 TGTGGCGGCAAGGCGGACCAACCGGATCCCGGATCAAGATCGGCGATCGGCGCTCGG 835

Qy 124 gggaaactcttgatgagcagtcgagatgagtaagaaggagagcgtctctccctgcccagcgtca 183
Db 836 TGGAAAGCTCAAGTTCAAGGGCTACGAGGACCGCTACACCATCGCGCGCGCTACTGGC 895

Qy 184 cagcgagatcagatcagcagcgccgggtccactctcgtctgctggcagaagtaactcgcgagccc 243
Db 896 CGCGCAGTTCGCGATCATGATGACGCGGAGACCCCTGGAACCGAAGCAGATGCTGTCCACCC 955

Qy 244 aaggcaagcctatggcagcagaagttcggcgagtgctcctcgcggaagcagctatca 303
Db 956 GCGGCGATACCGTATAGACCCAGACCCAGACCTACCCACCGGAACCGCGCGTGGCGGCGATCATCG 1015

Qy 304 agggatcagcgtcacctggtgagagcatca 333
Db 1016 CCTCCGACGACCCCGAGTTCATCGTCA 1045

RESULT 6
US-08-340-011-1
; Sequence 1, Application US/08340011
; Patent No. 5776755
; GENERAL INFORMATION:
; APPLICANT: Alitalo, et al.
; TITLE OF INVENTION: FLT4, A NOVEL RECEPTOR TYROSIN KINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,011
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/959,951
; FILING DATE: 09-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Thomas C.
; REGISTRATION NUMBER: 36,989
; REFERENCE/DOCKET NUMBER: 32267
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20...3916
; US-08-340-011-1

Query Match          10.7%; Score 41.2; DB 2; Length 4195;
Best Local Similarity 46.3%; Pred. No. 0.023;
Matches 136; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 51 ggaatcgcggggcaatcgccgacgagctgaggaagggagggagaaacggtccggtcgtc 110
Db 2075 GACCTTCCTGGTGAACGCTGAGCGCATCGCTGGAGATGCAGTCTTGGTGGCGGAGCGGCAC 2134

Qy 111 tcttcctctctctggaaactctttgatgagcagtcgagtcgagatgagtaacaaggagagcgtcctc 170
Db 2135 GCGCCAGCATCGTGTGGTACAAAGCAGGAGGCTGCTGGAGGAAAAGTCTGGAGTGCAC 2194

Qy 171 cctcgcgacgtcacagcagcaggtatcagatcagtcgagcggcggtccactctcgtctggcagaag 230
Db 2195 TTGGCGGACTCCACAGCAAGCTGAGCATCCAGCGCTGCGCGAGGAGATCGCGGACGC 2254

Qy 231 taagtctgagcccaaggcaagccattggtcagatcgacacaaagttcggcgagtgctcctgcc 290
Db 2255 TATCTGTGAGCGGTGTGCANCCCAAGGCTCGCTCACTCTCCGCCAGCGGTGGCCGTG 2314

Qy 291 gggacgatctacaaggagtagcgcatcacccgtggagagacatcttgcaactgccc 344
Db 2315 GAAGGCTCCGAGGATAAAGGCGAGCATGGAGATCGTGATCCTTGTGCGTACCGGC 2368

RESULT 7
US-08-901-710-1
; Sequence 1, Application US/08901710
; Patent No. 6107046
; GENERAL INFORMATION:
; APPLICANT: Alitalo, Kari
; APPLICANT: Aprelikova, Olga
; APPLICANT: Pajusola, Katri
; APPLICANT: Armstrong, Elina
; APPLICANT: Korhonen, Jaana
; APPLICANT: Kaipainen, Arja
; APPLICANT: Matikainen, Marja-Terttu
; TITLE OF INVENTION: FLT4, A RECEPTOR TYROSINE KINASE, AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```


Query Match	10.7%;	Score 41.2;	DB 6;	Length 9108;
Best Local Similarity	46.3%;	Pred. No. 0.03;		
Matches 136;	Conservative 0;	Mismatches 158;	Indels 0;	Gaps 0;
51	qagatcgcggaatcgccgcacagctgaagaagaggggaaacacagtcgcgcctc	110		
3018	GACCTCCTGGTGAACGTGAGCAGCTGCGTGGAGATGCATGTTGGTGGCCGGAGCGCAC	3077		
111	tgcttcgtctctcggaactctttgatgacgctcggtatgatacaagagagcgtctc	170		
3078	GGCCCCAGCATCGTGTGGTACAAAGACGAGAGGCTGCTGGAGGAAAAGTCTGGAGTCGAC	3137		
171	ctgtccgaagctcacagcagagatcagcatcgagccgggtccaactctcggtgcgagaag	230		
3138	TTGGCGGATCCCAACACGAAGCTGAGACTCCAGCGCGTGCAGAGGAGATGTCGGGAGCGC	3197		
231	tacgttcgagcccaagcgaaagccatttgcatctgcacaagtctccgcgcagtgctcctgcc	290		

Db 3198 TATCTGTCCAGCGTGTGCAAGCCCAAGGCGTGCCTCAACTCCTCCGCGAGCGTGGCGGTG 3257
Qy 291 gggagcattcaaggagctacgcatcacccctgagagcattgcaactgcc 344
Db 3258 GAAGCTCCGAGGATAGGCGCAGCATGAGATCGTGATCTTGTGCGTACCGGC 3311

RESULT 13

US-09-320-878-23/c
; Sequence 23, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880.
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 1693
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-23

Query Match 10.6%; Score 41; DB 5; Length 1693;
Best Local Similarity 52.0%; Pred. No. 0.02;
Matches 92; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
Qy 158 ggaagagcgtctccctgccagctcacagcagagatcagcatcgaggccgggtccactct 217
Db 1003 GAAGAAGCGTCTCGCGCGCGGTCTCGCTCGCGCTGCGAGGTCCAGGTGCGTACGCGAT 944
Qy 218 cggctggcagaagtacgtcgagcccaaggccattggcatgacacagttcgagcgc 277
Db 943 CGGAAGCCGAGCCCGTGGCGCTTCCTGCGAGACCTTCGGACCCACCAAGATCGAGAC 884
Qy 278 gagtgtctctccgggagcagctactacaaggagtagcggcatcacctggagagcatcat 334
Db 883 CGAGAAGATCGAAGACGCCATCGGCGAGGCTTCGACCTTCGCCCGCGCGCATCAT 827

RESULT 14

US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: HORLICK, Robert A.
; APPLICANT: Robbins, Bassam B.
; APPLICANT: Dama, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06

; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match 9.4%; Score 36.2; DB 4; Length 1931;
Best Local Similarity 45.6%; Pred. No. 0.4;
Matches 128; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
Qy 27 ggggtgggcaccggtccgagctcgagatcgagcgaatcgccgagcagcagctgaggaag 86
Db 505 GGGGAGGACGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGAG 446
Qy 87 ggggggaagacggtccgctcgtctctcgtctcctctggaactctttgatgagcagtcg 146
Db 445 GACGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGG 386
Qy 147 gatgagtcacaagagagcgtccctccctcgagcgtcacagcagcagcagcagcagcagc 206
Db 385 GAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGGGGAGGACGAGGAC 326
Qy 207 ggggtccactctcggtcgagcagagtcggtcgagcagcagcagcagcagcagcagcagc 266
Db 325 GGGGAGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGAC 266
Qy 267 aagttcgccgagtgctcctcctcgccggagcagctacacagga 307
Db 265 AAGGTGGCACCCAGGAAACGTCGGTTACGTTGAACCTGCA 225

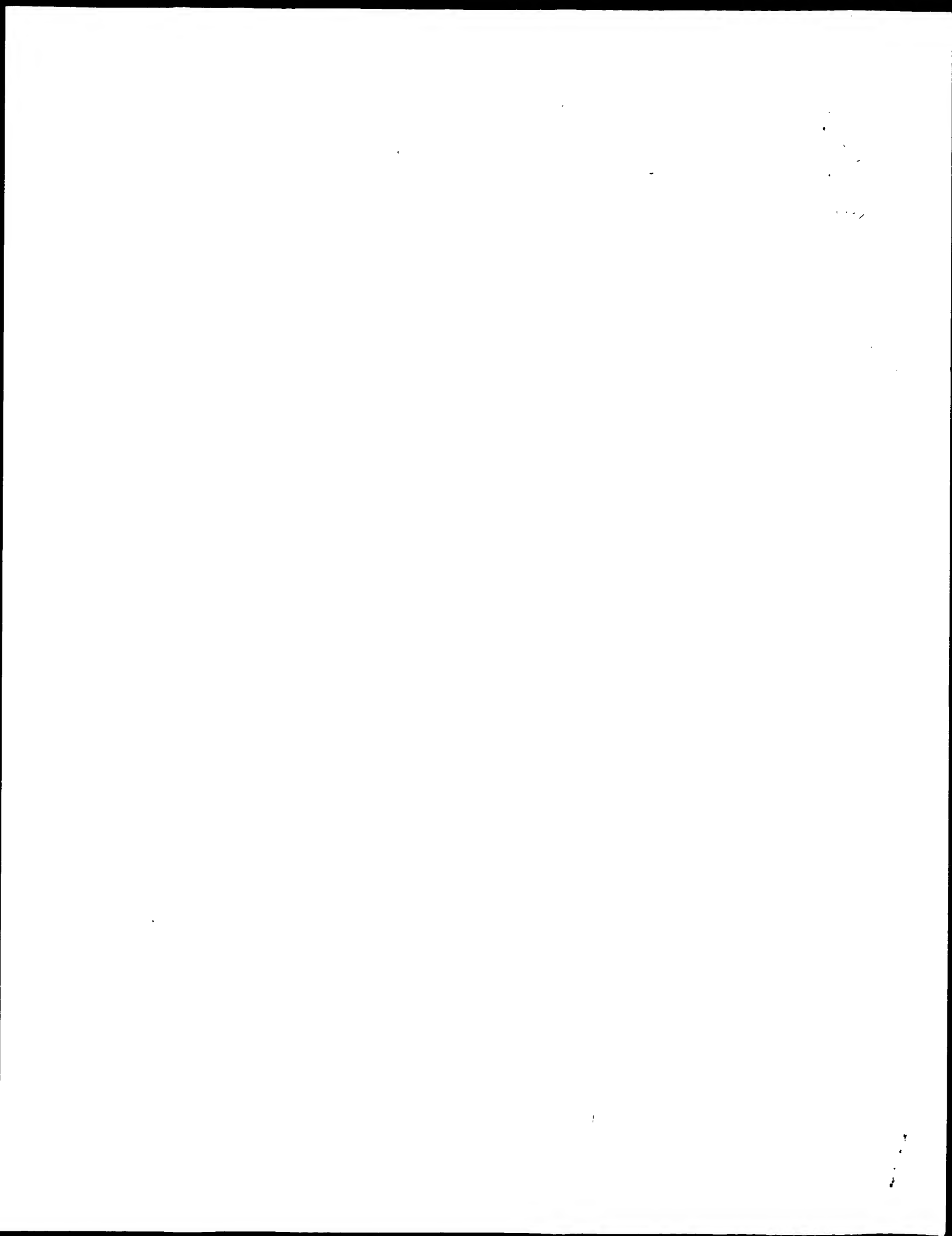
RESULT 15

US-08-106-981-5
; Sequence 5, Application US/08106981
; Patent No. 5731419
; GENERAL INFORMATION:
; APPLICANT: SARHAN, Fathey
; APPLICANT: HOUEDE, Mario
; APPLICANT: LALIBERT, Jean-Fran ois
; TITLE OF INVENTION: DNA MOLECULES ENCODING FREEZING
; TITLE OF INVENTION: TOLERANCE PROTEINS IN GRAMINEAE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/106,981
; FILING DATE: 16-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 163-23
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1136
; TYPE: Nucleic acid

; STRANDEDNESS: Double
; TOPOLOGY: Linear
; MOLECULE TYPE: CDNA to mRNA
; ORIGINAL SOURCE: Triticum Aestivum L.
US-08-106-981-5

Query Match 9.2%; Score 35.6; DB 2; Length 1136;
Best Local Similarity 51.2%; Pred. No. 0.49; 79; Indels 0; Gaps 0;
Matches 83; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
Qy 4 ccggcaccagcctgacatcattgggttgggcacccggctccgagctggagatcgcgggca 63
Db 130 CGCGCGAGCAGGTGGAGGTGACGACAGGGGCTCTCTGGCAACCTCTCTGGCAAGA 189
Qy 64 atcgccgcgacgctgaggaagagggggaagacgggtccgctgctctctctct 123
Db 190 AGGCTGAGGAGGACAAAGGAGAAGGAGGAGCTGGTCACCGGATGGAGAAGTCTCCG 249
Qy 124 gggaactctttagatgagcagtcgagatagatcaagagagcg 165
Db 250 TGGAGAGCCCGAGGTCAAGAGGAGGAGGACACGAGGATGGCG 291

Search completed: November 4, 2000, 13:34:17
Job time: 16693 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:33:25 ; Search time 4075.18 Seconds
(without alignments)
338.673 Million cell updates/sec

Title: US-09-300-482-225
Perfect score: 316
Sequence: 1 gataagtgccacactgag.....gaatggaggagactctcga 316

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl1.*
10: gb_pr2.*
11: gb_pr3.*
12: em_fun.*
13: em_hum1.*
14: em_hum2.*
15: em_in.*
16: em_om.*
17: em_or.*
18: em_ov.*
19: em_pat.*
20: em_ph.*
21: em_pl.*
22: em_ro.*
23: em_sts.*
24: em_sy.*
25: em_un.*
26: em_vi.*
27: gb_htg1.*
28: gb_htg2.*
29: gb_inl.*
30: gb_in2.*
31: em_ba1.*
32: em_ba2.*
33: em_hum3.*
34: em_hum4.*
35: gb_pr4.*
36: gb_htg3.*
37: gb_htg4.*
38: gb_htg5.*
39: gb_htg6.*
40: gb_htg7.*
41: em_htg1.*
42: em_htg2.*
43: em_htg3.*

44: em_hum5.*
45: gb_pl3.*
46: gb_pr5.*
47: gb_htg8.*
48: gb_htg9.*
49: gb_htg10.*
50: gb_htg11.*
51: gb_htg12.*
52: gb_htg13.*
53: gb_htg14.*
54: gb_in3.*
55: gb_htg15.*
56: gb_htg16.*
57: gb_htg17.*
58: em_htg4.*
59: em_htg5.*
60: em_htg6.*
61: em_htg7.*
62: em_hum6.*
63: gb_htg18.*
64: gb_htg19.*
65: gb_ba3.*
66: em_htg8.*
67: em_htg9.*
68: em_htg10.*
69: em_htg11.*
70: em_htg12.*
71: em_htg13.*
72: em_htg14.*
73: em_htg15.*
74: em_htg16.*
75: em_htg17.*
76: em_htg18.*
77: em_htg19.*
78: em_htg20.*
79: em_htg21.*
80: em_htg22.*
81: em_htg23.*
82: gb_pr6.*
83: gb_pr7.*
84: gb_htg20.*
85: gb_htg21.*
86: gb_htg22.*
87: gb_htg23.*
88: gb_ro.*
89: gb_sts1.*
90: gb_sts2.*
91: gb_sy.*
92: gb_un.*
93: gb_vil.*
94: gb_vil2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	160.8	50.9	930	7	AF189365	AF189365 Oryza sat
c 2	78.4	24.8	89154	8	ATAC010797	AC010797 Arabidops
3	69.8	22.1	67712	8	ATAC022355	AC022355 Arabidops
4	68.4	21.6	5348	45	S77133	S77133 GRF1-genera
5	63	19.9	720	8	CNS01918	AL111059 Botrytis
c 6	60.6	19.2	2913	45	ZMU29162	U29162 Zea mays cl
c 7	57.8	18.3	424	89	G37533	G37533 SHGC-57943
8	54.8	17.3	1574	82	HSJA74326	AJ224326 Homo sapi
9	53.8	17.0	10304	7	AF166527	AF166527 Zea mays
c 10	52.2	16.5	3846	7	AB031012	AB031012 Zea mays
11	51.4	16.3	26292	40	AC019942	AC019942 Drosophila
c 12	51.4	16.3	27521	30	CELF08F8	U28991 Caenorhabdi

C 13 51.4 16.3 50089 29 AC005641
 C 14 51.4 16.3 125150 27 AC005450
 C 15 51.4 16.3 131899 27 AC008258
 C 16 51.4 16.3 259764 29 AE003840
 C 17 50.4 15.9 122925 85 AL360001
 C 18 49.8 15.8 119907 27 AC008259
 C 19 47.6 15.1 911 45 SCPOS18
 C 20 47.6 15.1 1328 45 SCYJL121C
 C 21 47.6 15.1 39729 45 SPAC31G5
 C 22 47 14.9 5537 83 ZM2MM1
 C 23 46.4 14.7 1127 83 HUMARS1H
 C 24 45.6 14.4 182756 7 AC007789
 C 25 44.6 14.1 82101 10 AC005894
 C 26 44.6 14.1 126138 82 HSAY18000
 C 27 44.6 14.1 149308 9 AC005527
 C 28 44.6 14.1 153804 27 AC006280
 C 29 44.6 14.1 196149 27 AC004709
 C 30 44.6 14.1 318488 9 AC005529
 C 31 44.2 14.0 1050 7 AF047444
 C 32 44.2 14.0 135209 46 HS42616
 C 33 44 13.9 7753 45 ZMAYSPG
 C 34 43.4 13.7 8291 45 ZM057899
 C 35 42.4 13.4 7753 45 ZMAYSPG
 C 36 41.8 13.2 6414 45 ZMGPAL
 C 37 41.8 13.2 6414 45 STPEPELMR
 C 38 41.2 13.0 1055 45
 C 39 39.6 12.5 37262 29 AC009805
 C 40 39.6 12.5 39475 29 AC016024
 C 41 39 12.3 182756 7 AC007789
 C 42 38 12.0 1307 2 MAGHPS
 C 43 37.8 12.0 5537 45 ZM2MM1
 C 44 37.6 11.9 139510 39 AC016779
 C 45 37.2 11.8 10304 7 AF166527

AC005641 Drosophil
 AC005450 Drosophil
 AC008258 Drosophil
 AE003840 Drosophil
 AL360001 Homo sapi
 AC008259 Drosophil
 X83571 S. cerevisia
 Z49396 S. cerevisia
 Z98979 S. pombe chr
 X81199 Z.mays ZMM1
 J04742 Human auton
 AC007789 Oryza sat
 AC005894 Homo sapi
 Y18000 Homo sapien
 AC005527 Homo sapi
 AC006280 Plasmodiu
 AC004709 Plasmodiu
 AC005529 Homo sapi
 AF047444 Oryza sat
 AL020997 Human DNA
 Z11879 Z.mays P ge
 U57899 Zea mays re
 Z11879 Z.mays P ge
 M18976 Zea mays ch
 X15408 Maize Gpal
 Z50098 S. tuberosum
 AC009805 Leishmani
 AC016024 Leishmani
 AC007789 Oryza sat
 D64136 Methylomona
 X81199 Z.mays ZMM1
 AC016779 Oryza sat
 AF166527 Zea mays

RESULT 1
 AF189365
 LOCUS Oryza sativa 930 bp mRNA PLN 10-JAN-2000
 DEFINITION Oryza sativa D-ribulose-5-phosphate 3-epimerase mRNA, complete cds.
 ACCESSION AF189365
 VERSION AF189365.1 GI:6007802
 KEYWORDS
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
 Kopriya, S., Koprivova, A. and Suss, K. H.
 Identification, cloning, and properties of cytosolic
 D-ribulose-5-phosphate 3-epimerase from higher plants
 J. Biol. Chem. 275 (2), 1294-1299 (2000)
 20092904
 2 (bases 1 to 930)
 Kopriya, S., Koprivova, A. and Suss, K. H.
 Direct Submission
 Submitted (20-SEP-1999) Institute of Forest Biology and Tree
 Physiology, Am Flughafen 17, Freiburg im Breisgau 79085, Germany
 Location/Qualifiers
 1..930
 /organism="Oryza sativa"
 /db_xref="taxon:4530"
 35..721
 /EC_number="5.1.3.1"
 /note="cytosolic isoform"
 /codon_start=1
 /product="D-ribulose-5-phosphate 3-epimerase"
 /protein_id="AA01048.1"
 /db_xref="GI:6007803"
 /translation="MAAARAAKIAPSMLSSDFANLAAFAADRVRLGADWLHMDIMDGH
 FVPNLTIAGPVISQSRKTKAYLDCHLMTNPDSYVEFLAKAGASGFTFIEVSRDNK

Query Match 50.9%; Score 160.8; DB 7; Length 930;
 Best Local Similarity 83.2%; Pred. No. 1.6e-41;
 Matches 183; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
 QY 1 gataaggtgcgcacactgagaaagaagcaccctcccttcacatagaggttgatgggtg 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 518 GAAAGAGTGCCTGCCTCAGAAAGAGTACCATCCCTTGACATCGAGGTGATGGTGGT 577
 QY 61 ctatgtctcttaacacatagagcgtggccgcacatctgtgtgggccaattgcctgcctgga 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 578 CTGGGTCTTCACCATGATGATGTGGCTGCATCTGTGGGGCTAACTGATTTGCTCTGGA 637
 QY 121 agctctatatgtggcgtgcgcagccagcagcagcagcagcagcagcagcagcagcagc 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 638 ACTCAATATTGGAGCAGCTGAACAGAGAGGTGCATATCGGCACCTAAGGAAGAGCGTT 697
 QY 181 gagggctctcagaacaaactgatttgggtgttctgct 220
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 698 GAAGGATCTCAGAACAAAAGCTGATCTGTGTGGTTTGTACT 737

RESULT 2
 ATAC010797/c
 LOCUS Arabidopsis thaliana chromosome III BAC F28J7 genomic sequence,
 DEFINITION complete sequence.
 ACCESSION AC010797
 VERSION AC010797.2 GI:6091712
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsids.
 Lin, X., Kaul, S., Town, C. D., Benito, M., Creasy, T. H., Haas, B.,
 Ronning, C. M., Koo, H., Fujii, C. Y., Utterback, T. R., Barnstead, M. E.,
 Bowman, C. L., White, O., Nierman, W. C. and Fraser, C. M.
 Arabidopsis thaliana chromosome III BAC F28J7 genomic sequence
 Unpublished
 2 (bases 1 to 89154)
 Lin, X. and Kaul, S.
 Direct Submission
 Submitted (23-SEP-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org
 3 (bases 1 to 89154)
 Lin, X.
 Direct Submission
 Submitted (21-OCT-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 4 (bases 1 to 89154)
 Lin, X.
 Direct Submission
 Submitted (09-DEC-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Oct 21, 1999 this sequence version replaced gi:5919288.
 Address all correspondence to:
 Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlin@tigr.org
 BAC clone F28J7 is from Arabidopsis chromosome III and is near the
 molecular marker ILRI.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://www.cds.stanford.edu/~chris/genscanw.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tadb/at/at.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

FEATURES

source

1. 89154

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="III"

/map="ILRI"

/clone="F28J7"

1. 28249

/note="Overlap with BAC clone F4P13

(AC009325:77296..105543)."

complement(784..815)

/note="exon predicted by xgrail, quality

marginal_shadowexon"

join(<1449..1750,2015..2051,2153..2389,2484..2531,

2614..3264)

/gene="F28J7.1"

<1449..3264

/note="F28J7.1"

/gene="F28J7.1"

/note="F4P13.23; similar to putative protein GB:CAA22987

[Arabidopsis thaliana]

join(1449..1750,2015..2051,2153..2389,2484..2531,

2614..3075)

/gene="F28J7.1"

/note="unknown protein"

/codon_start=1

/db_xref="GI:6091713"

/translation="NGVYTSVAAKFAFPSPPSYKVVTDELTLGLLLSPFFPHRENV

EIVKLRARGEIVGMVYRHPMATSTLLYSHGNAIDLGOMYELFIELSHKYNLMGY

DYSGYQSTGPKPSHNTYADIEAVYKCLEETFGSKQEGVILYGVSGHPTLDASRL

POLRAVILHSILSLRVYVSVKTYWFDIKYKIDKIPYVDCPVLIHGTSDRWDCS

HGQLWELCKDKYEPPLWVKGNGHCDLHYPEYIPHLKFIATYVERLPCPRMSDQSER

VNDAPPKRSMDRRVKPQSTERREREKPKPSQKSMSSSSKLIKISFDQLDRSRSDVC

HEKTRKSDQIERGKSVDRDLRVSE"

<3677..4136

/gene="F28J7.2"

<3677..4136

/gene="F28J7.2"

/note="F4P13.24"

3677..4087

/gene="F28J7.2"

/note="unknown protein"

/codon_start=1

/db_xref="GI:6091714"

/translation="MARLFVYVALLALVGTFAADAPSNAPTASTPKFTKAPAAAP

KSSAAAPKASPAVEETPEDDYSAASPSASAPTVSSPAPTPEADGPSDDGSSD

GPAAEPSKSAATNVKLSTAGIVAAAGFFIFSL"

4330..4401

/gene="F28J7.3"

/product="tRNA-Pro"

gene

mRNA

gene

CDS

misc_feature

misc_feature

mRNA

gene

CDS

mRNA

gene

CDS

trRNA

/anticodon="(pos:4362..4364,aa:Pro)

4330..4401

/gene="F28J7.3"

complement(join(<4853..4932,5366..5690,5785..6288,

6384..>6434))

/gene="F28J7.4"

complement(<4853..>6434)

/gene="F28J7.4"

/note="F4P13.26; predicted by genefinder"

complement(join(4853..4932,5366..5690,5785..6288,

6384..6434))

/gene="F28J7.4"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAF03427.1"

/db_xref="GI:6091715"

/translation="MGESACLMQFYSYAAPQDGLGALGQSVSGFRMSEKLDEKWS

AFSTQNPYVAEARYSKPGSAQKAFEAHYKVAARAKAAAEALLQQOIKPKOP

EKSPQVEDINGEKKDMVSKPDLEIPRGLDVSQAKLLAEKGRFSNRQSEKEN

CGMAESEINGRTAEQVDEKPIKLSVKKSKDSQPKSSTPRVSKYNSERTPSKRS

NNKSSSYTTPAKEFNRLYSIIIRKIDGSRASKPKRDKTLPRTSPKNTVAFFKLCT

QPSAYALVMIAGVCKGNEARGEIQGNGDTEPEGGEGEGEGRE"

complement(join(<7108..7719,7796..8075,8138..8302,

8384..8859,8940..9140,9249..9492,9578..9748,

10009..>10304))

/gene="F28J7.5"

complement(<7108..>10304)

/gene="F28J7.5"

/note="F4P13.27"

complement(join(7108..7719,7796..8075,8138..8302,

8384..8859,8940..9140,9249..9492,9578..9748,10009..10304))

/gene="F28J7.5"

/note="unknown protein"

/codon_start=1

/protein_id="AAF03428.1"

/db_xref="GI:6091716"

/translation="MRWDLITAIVAALVSVLADESGQMAPYRIHTLFSVEQNYFOW

QTVGLMHSFLKSGQPGPTIRLLSCTDDQKTYIRGMNLAPTPEVPSWRHPKTDGWPFA

INKPYGVLYWLQHSSEAKHVDWVILDDADMIIRGPIIPWELGAERPRFAAHYGLVG

CUNLILVRLHTKHPCLDKVGLLAHIDDLRLVLAFLSKTEDVRODTAHTWNTLGD

LYGKWISEMXYSGFAEGAKHKINDDLMIYGYVPRGVEPVLHMYGLPFSIGNM

SFTKIDHEDNIVDCNRLFFPEPPREVKIMEPPDKRRGLILSLECMNTLNEGLIL

RHAENGCPKAWTKYLSFLKTFMELTRPKLAPGSVHILPDQHEPPIDEFKGTYV

KIHTLFTECTYFDWQVGFHSGFRQSGQPGNTRLLSCTDEALKNYKGDLPATHY

VPSMRHPLTGDWYPAINKPAAVVGLHHTNIDAEYVVIDADMILRGPIITPEFKAA

RGPRVTPESPLKPSLFLSYLIGDNDLRLHTRNPEACDKVGGVILIHIEDLRK

FAMYVLLKTOVRADKHYGKELTDIVESGWISEMISYSGFAELRLHRSINKEIMI

YGYVPEGADYRVHYGLEKFGVGNWDFRANRNTDLINKWAKFPDPPPSAVHOT

DNDLRQDLLEGGKLINEALFHHRRNCPEPSESTEKISVRKVGNIETKQOG

SDTKESGSSSESGRFTSLKLWLVIALWISGVFLVMLLVSTRRGRTTRGKGYR

NKRRTSYNTGFLDTR"

complement(10410..10481)

/note="exon predicted by xgrail, quality good"

10571..10626

/note="exon predicted by xgrail, quality

marginal_shadowexon"

complement(10844..10930)

/note="exon predicted by xgrail, quality

marginal_shadowexon"

complement(11233..11268)

/note="exon predicted by xgrail, quality

marginal_shadowexon"

<11324..>11587

/gene="F28J7.6"

/note="F28J7.6"

/note="F4P13.28; predicted by genscan"

11324..11587

/gene="F28J7.6"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAF03429.1"

	BASE COUNT	1325 a	1169 c	1203 g	1651 t	
	ORIGIN					
	Query Match	21.6%	Score 68.4;	DB 45;	Length 5348;	
	Best Local Similarity	87.2%;	Pred. No. 3.2e-11;	Mismatches 11;	Indels 0;	Gaps 0;
	Matches	75;	Conservative	0;		
Qy	225	agtaactccctcggttttttttatttcgctgcgttttagttcaaacatgaactagcggacg	284			
Db	3413	AATACTCCCTCGGTGTTTATTATTCACGTTTATAGTAATAAATAACTAGCGGGC	3472			
Qy	285	actgatattcgagaatggaggagta	310			
Db	3473	ACAATATTCGAGAATGGAGGTAGTA	3498			
	RESULT	5				
	CNS01918	720 bp	mRNA	PLN	02-SEP-1999	
	LOCUS	Botrytis cinerea strain T4 cdna library under conditions of nitrogen deprivation.				
	DEFINITION	AL111059				
	ACCESSION	AL111059.1	GI:5825679			
	VERSION	cdna library; nitrogen deprivation.				
	KEYWORDS	Botryotinia fuckeliana				
	SOURCE	Botryotinia fuckeliana				
	ORGANISM	Eukaryota; Fungi; Ascomycota; Euascomycetes; Discomycetes; Leotiiales; Sclerotiniaceae; Botryotinia.				
	REFERENCE	1 (bases 1 to 720)				
	AUTHORS	Bitton,F., Lewis,C., Fortini,D., Pradier,J.M. and Brygoo,Y.				
	TITLE	Direct Submission				
	JOURNAL	Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr, 78026 Versailles, France				
	REFERENCE	2 (bases 1 to 720)				
	AUTHORS	Genoscope.				
	TITLE	Direct Submission				
	JOURNAL	Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage : CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
	COMMENT	- Web : www.genoscope.cns.fr) The cdna library to be analyzed within the framework of this project was created using a Botrytis cinerea strain which was grown under conditions of nitrogen deprivation, which is the normal situation for B. cinerea during its development on its host plant. The library was produced in an oriented direction, in the pBSII vector.				
	FEATURES	Location/Qualifiers				
	source	1..720				
	/organism="Botryotinia fuckeliana"					
	/strain="T4"					
	/db_xref="taxon:40559"					
	/note="Genoscope sequence ID : W62E091"					
	BASE COUNT	226 a	146 c	178 g	170 t	
	ORIGIN					
	Query Match	19.9%;	Score 63;	DB 8;	Length 720;	
	Best Local Similarity	65.0%;	Pred. No. 1.3e-09;	Mismatches 50;	Indels 0;	Gaps 0;
	Matches	93;	Conservative	0;		
Qy	3	taagtgcgcacactgagaagaagtacccttcoccttgacatagagtggtggtgtct	62			
Db	578	TAAGTGCCGAACACTCAGAAACGATACCAGACCTCAACATCGAAGTTGATGGTGATT	637			
Qy	63	aggtccttcaccatagacgtggccgcatctgctggggccaattgcctgcgtcgtaa	122			
Db	638	GGGTCTCGAACAAATGTATCAAGCGGAGAPGCTGGCGCAAAATGTGNTGGTCAAGCAG	697			
Qy	123	ctctatatattggcgtgcgcgacc	145			
Db	698	TGCTGTTTTGGAGCCAAGGACC	720			


```

Db 7668 AAGTGTACTTCCATTCCTTTTATTGTCGTTTATTCAAAATGAACTAGTAG 7727
QY 283 CGACTGATTCGAGAGTGGAGGAGTACTT 313
Db 7728 CGACAAATATTCGAGAACTGAGGTAGTGT 7758

RESULT 10
AB031012/c
LOCUS AB031012 3846 bp DNA PLN 12-JUL-2000
DEFINITION Zea mays ZmRR2 gene for response regulator 2, complete cds.
ACCESSION AB031012
VERSION AB031012.1 GI:6009896
KEYWORDS response regulator 2.
SOURCE Zea mays (sub_species:Golden Cross Bantam T51) DNA.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
Deji,A., Sakakibara,H., Ishida,Y., Yamada,S., Komari,T., Kubo,T.
and Sugiyama,T.
Genomic organization and transcriptional regulation of maize ZmRR1
and ZmRR2 encoding cytokinin-inducible response regulators
Biochim. Biophys. Acta 1492, 216-220 (2000)
REFERENCE 2 (bases 1 to 3846)
AUTHORS Sakakibara,H., Deji,A. and Sugiyama,T.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) to the DDBJ/EMBL/GenBank databases. Hitoshi
Sakakibara, Nagoya University, Graduate School of Bioagricultural
Sciences, Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan
(E-mail:sakakibagr.nagoya-u.ac.jp, Tel:81-52-789-4105,
Fax:81-52-789-4107)
FEATURES
source
1. .3846
/organism="Zea mays"
/sub_species="Golden Cross Bantam T51"
/db_xref="taxon:4577"
2842. .2849
join(3000..3186,3313..3599)
/gene="ZmRR2"
join(3000..3186,3313..3599)
/gene="ZmRR2"
/note="cytokinin-inducible"
/codon_start=1
/product="response regulator 2"
/protein_id="BA85113.1"
/db_xref="GI:6009897"
/translation="MAAAATATPSVAPESGDRKAVPPVDVDAVLELEKEHVLAVD
SSVDRAVIKLRSSRYRTTVDSATRALELLALGLVDPVNMIIIDYMPGMTGYELL
KHVSSALRAIPVIMSSENVTRISLCEGAEDFLKPKVPADVSLCSRR"
3187..3312
BASE COUNT 975 a 883 c 971 g 1017 t
ORIGIN

Query Match 16.5%; Score 52.2; DB 7; Length 3846;
Best Local Similarity 82.2%; Pred. No. 5.9e-06;
Matches 60; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 238 ttttttttttgcgcgttttagttcaacatgaactgcgagactgattcag 297
Db 1061 TCTTTTTTTTGTGCTGTTAGTTTAAATGAACATAGTACGACCAATATTCGAG 1002

QY 298 aatgaggaggagta 310
Db 1001 AACGGAGATAGTA 989

RESULT 11
AC019942
LOCUS AC019942 26292 bp DNA HTG 03-JAN-2000
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered

```

```

pieces.
AC019942
VERSION AC019942.1 GI:6664955
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 26292)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDW:10211468 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. .26292
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 7122 a 6111 c 5943 g 7116 t
ORIGIN

Query Match 16.3%; Score 51.4; DB 40; Length 26292;
Best Local Similarity 53.2%; Pred. No. 1.5e-05;
Matches 109; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 4 agagtgccacactgagaaagtagtaccctccctccctgacatagagattgattgctcta 63
Db 24474 AAGGTCAAGTGGCTGCGCGAAACTACCCCAACTGGACATCGAGGTGACGAGGTG 24533

QY 64 ggtcctcaaccatgacgtgcccgcattctgtggggccaattgcattgcctggaagc 123
Db 24534 GGACCAAGACTATACACTGCTGTGCCGAGCGCGACCAACATGATCTCTCGGAACC 24593

QY 124 tctatttggcgtgcgagaccagagccatcatctctgtctgaggaagagcgtcag 183
Db 24594 GCGGTGGTGGCGCTCCCATCAGTCGCAGGTCAACAGGAGTGGCGGATGTGTGCAC 24653

QY 184 ggcctcagaacacaaactgattt 208
Db 24654 AGCTACCTCAATAGATTCATTT 24678

RESULT 12
CELLF08F8/c
LOCUS CELLF08F8 27521 bp DNA INV 15-JUN-1995
DEFINITION Caenorhabditis elegans cosmid F08F8.
ACCESSION U28991
VERSION U28991.1 GI:861364
KEYWORDS
SOURCE Caenorhabditis elegans strain-Bristol N2.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 27521)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Smith,A.,
Saunders,D., Showkeen,R., Smaiden,N., Smith,A., Sonhammer,E.,
Staden,K., Waterston,J., Thierri-Mieg,J., Thomas,K., Vaudin,M.,
Wilkinson-Sproat,J. and Wohlman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

```


Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE AUTHORS

1 (bases 1 to 125150)
Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenikoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.I. and
Rubin,G.M.

Sequencing of Drosophila melanogaster

TITLE JOURNAL REFERENCE AUTHORS

2 (bases 1 to 125150)
Celnikier,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,
Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
Zieran,L.I. and Kimmel,B.E.

TITLE JOURNAL

Direct Submission
Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT

On Jul 30, 1999 this sequence version replaced gi:5597051.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bagp@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases. Pl library location:
7-33.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 610: contig of 610 bp in length

* 611 690: gap of unknown length

* 691 1308: contig of 618 bp in length

* 1309 1388: gap of unknown length

* 1389 1943: contig of 554 bp in length

* 1943 2022: gap of unknown length

* 2022 2583: contig of 561 bp in length

* 2583 2663: gap of unknown length

* 2663 3865: contig of 1202 bp in length

* 3865 3945: gap of unknown length

* 3945 4977: contig of 1032 bp in length

* 4977 5057: gap of unknown length

* 5057 5719: contig of 662 bp in length

* 5719 5799: gap of unknown length

* 5799 6451: contig of 652 bp in length

* 6451 6531: gap of unknown length

* 6531 7403: contig of 872 bp in length

* 7403 7483: gap of unknown length

* 7483 8621: contig of 1138 bp in length

* 8621 8701: gap of unknown length

* 8701 9614: contig of 913 bp in length

* 9614 9694: gap of unknown length

* 9694 10744: contig of 1050 bp in length

* 10744 10824: gap of unknown length

* 10824 11457: contig of 633 bp in length

* 11457 11538: gap of unknown length

* 11538 12289: contig of 752 bp in length

* 12289 12369: gap of unknown length

* 12369 14469: contig of 2100 bp in length

* 14469 14549: gap of unknown length

* 14549 15325: contig of 776 bp in length

* 14550

15326 15405: gap of unknown length
15405 16338: contig of 933 bp in length
16338 16418: gap of unknown length
16418 17558: contig of 1138 bp in length
17558 17636: gap of unknown length
17636 20200: contig of 2564 bp in length
20200 20280: gap of unknown length
20280 21724: contig of 1444 bp in length
21724 21804: gap of unknown length
21804 22429: contig of 625 bp in length
22429 22509: gap of unknown length
22509 23332: contig of 823 bp in length
23332 23412: gap of unknown length
23412 24354: contig of 942 bp in length
24354 24434: gap of unknown length
24434 26644: contig of 2210 bp in length
26644 26724: gap of unknown length
26724 28963: contig of 2239 bp in length
28963 29043: gap of unknown length
29043 37172: contig of 8129 bp in length
37172 37252: gap of unknown length
37252 110032: contig of 72780 bp in length
110032 110112: gap of unknown length
110112 110721: contig of 609 bp in length
110721 110801: gap of unknown length
110801 111445: contig of 644 bp in length
111445 111525: gap of unknown length
111525 112124: contig of 599 bp in length
112124 112204: gap of unknown length
112204 112817: contig of 613 bp in length
112817 112897: gap of unknown length
112897 113527: contig of 630 bp in length
113527 113607: gap of unknown length
113607 114212: contig of 605 bp in length
114212 114292: gap of unknown length
114292 114906: contig of 614 bp in length
114906 114986: gap of unknown length
114986 115598: contig of 612 bp in length
115598 115678: gap of unknown length
115678 116267: contig of 589 bp in length
116267 116347: gap of unknown length
116347 116981: contig of 634 bp in length
116981 117061: gap of unknown length
117061 117399: contig of 338 bp in length
117399 117400: gap of unknown length
117400 118110: contig of 631 bp in length
118110 118190: gap of unknown length
118190 118336: contig of 646 bp in length
118336 118916: gap of unknown length
118916 119535: contig of 619 bp in length
119535 119615: gap of unknown length
119615 120248: contig of 633 bp in length
120248 120329: gap of unknown length
120329 120951: contig of 623 bp in length
120951 121031: gap of unknown length
121031 121644: contig of 613 bp in length
121644 121724: gap of unknown length
121724 122340: contig of 616 bp in length
122340 122420: gap of unknown length
122420 123036: contig of 616 bp in length
123036 123116: gap of unknown length
123116 123756: contig of 640 bp in length
123756 123836: gap of unknown length
123836 124448: contig of 612 bp in length
124448 124529: gap of unknown length
124529 125150: contig of 622 bp in length.

FEATURES source

Location/Qualifiers
1. .125150
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/clone="DS00609 (D364)"
/chromosome="2"

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (30-JUL-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jan 31, 2000 this sequence version replaced gi:5670394.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgpe@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 119 contigs. The true order of the pieces

```
* 34251 34330: gap of unknown length
* 34331 35143: contig of 813 bp in length
* 35144 35223: gap of unknown length
* 35224 36031: contig of 808 bp in length
* 36032 36111: gap of unknown length
* 36112 37040: contig of 929 bp in length
* 37041 37120: gap of unknown length
* 37121 37779: contig of 659 bp in length
* 37780 37859: gap of unknown length
* 37860 38785: contig of 926 bp in length
* 38786 38865: gap of unknown length
* 38866 39640: contig of 775 bp in length
* 39641 39720: gap of unknown length
* 39721 40670: contig of 950 bp in length
* 40671 40750: gap of unknown length
* 40751 41782: contig of 1032 bp in length
* 41783 41862: gap of unknown length
* 41863 43060: contig of 1198 bp in length
* 43061 43140: gap of unknown length
* 43141 44632: contig of 1492 bp in length
* 44633 44712: gap of unknown length
* 44713 45546: contig of 834 bp in length
* 45547 45626: gap of unknown length
* 45627 46916: contig of 1290 bp in length
* 46917 46996: gap of unknown length
* 46997 48619: contig of 1623 bp in length
* 48620 48699: gap of unknown length
* 48700 49921: contig of 1222 bp in length
* 49922 50001: gap of unknown length
* 50002 51409: contig of 1408 bp in length
* 51410 51489: gap of unknown length
* 51490 52306: contig of 817 bp in length
* 52307 52386: gap of unknown length
* 52387 53273: contig of 887 bp in length
* 53274 53353: gap of unknown length
* 53354 54488: contig of 1135 bp in length
* 54489 54568: gap of unknown length
* 54569 56096: contig of 1528 bp in length
* 56097 56176: gap of unknown length
* 56177 57247: contig of 1070 bp in length
* 57248 57326: gap of unknown length
* 57327 57936: contig of 610 bp in length
* 57937 58016: gap of unknown length
* 58017 59545: contig of 1529 bp in length
* 59546 59625: gap of unknown length
* 59626 60246: contig of 621 bp in length
* 60247 60326: gap of unknown length
* 60327 61305: contig of 979 bp in length
* 61306 61385: gap of unknown length
* 61386 62808: contig of 1423 bp in length
* 62809 62889: gap of unknown length
* 62890 64772: contig of 1884 bp in length
* 64773 64852: gap of unknown length
* 64853 66512: contig of 1660 bp in length
* 66513 66592: gap of unknown length
* 66593 67683: contig of 1091 bp in length
* 67684 67763: gap of unknown length
* 67764 69371: contig of 1608 bp in length
* 69372 69451: gap of unknown length
* 69452 70796: contig of 1345 bp in length
* 70797 70876: gap of unknown length
* 70877 72407: contig of 1531 bp in length
* 72408 72487: gap of unknown length
* 72488 73816: contig of 1329 bp in length
* 73817 73896: gap of unknown length
* 73897 75944: contig of 2048 bp in length
* 75945 76024: gap of unknown length
* 76025 78183: contig of 2159 bp in length
* 78184 78263: gap of unknown length
* 78264 80116: contig of 1853 bp in length
* 80117 80196: gap of unknown length
* 80197 81824: contig of 1628 bp in length
* 81825 81904: gap of unknown length
```

```
* 81905 84489: contig of 2585 bp in length
* 84490 84569: gap of unknown length
* 84570 86833: contig of 2264 bp in length
* 86834 86913: gap of unknown length
* 86914 89567: contig of 2654 bp in length
* 89568 89647: gap of unknown length
* 89648 92088: contig of 2441 bp in length
* 92089 92168: gap of unknown length
* 92169 94346: contig of 2178 bp in length
* 94347 94426: gap of unknown length

Query Match      16.3%; Score 51.4; DB 27; Length 131899;
Best Local Similarity 53.2%; Pred. No. 2e-05;
Matches 109; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 4 aaggtgcacactgagaaagaatgaccctcccttgacatagagggtgatggtccta 63
Db 46526 AAGGTCAAAGTGGCTGCGCGAAACTACCCCAACCTGGACATCGAGGTGGAGGTGTG 46467

QY 64 ggtccttcaaccatagacgtggccgcatctgctggggccaattgcatcgctggaagc 123
Db 46466 GGACCCCAAGACTATACACTGTGTGCGCGAGCGGAGCCACACATGATCGTCTCGGGAACC 46407

QY 124 tctatatttggcgtgcggaccaggagccatcatatctgtctgaggaagagcgtcgag 183
Db 46406 GCGGTGGTGGCGCCTCCGATCAGTCGCGAGGTCTCATCAAGGAGTTCGCGGATGTGGTGAC 46347

QY 184 ggctctcagaacaaactgatttt 208
Db 46346 AGCTACCTCAAATAGAGATTTCATTT 46322
```

Search completed: November 4, 2000, 13:34:27
Job time: 17116 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:44 ; Search time 320.8 Seconds
(without alignments)
351.306 Million cell updates/sec

Title: US-09-300-482-619

Perfect score: 300

Sequence: 1 gattattgacaacacgat.....ttcagaaatgtctgtgtgg 300

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36:*

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.2	12.1	10820	20	X20527
C 2	32.4	10.8	543	18	V75242
C 3	31.6	10.5	32768	20	X13336
C 4	30.8	10.3	579	20	X87836
C 5	30.8	10.3	1004	20	X13479
C 6	30.6	10.2	1051	19	V27386
C 7	30.6	10.2	13206	19	V52166
C 8	30.6	10.2	22243	18	V74475
C 9	29.8	9.9	1062	20	Z39041
C 10	29.8	9.9	1190	19	V61936
C 11	29.8	9.9	1308	20	X25510
C 12	29.8	9.9	1373	19	V52969

13	29.8	9.9	1683	21	Z99543
14	29.8	9.9	1750	20	V74136
15	29.8	9.9	1996	20	X25508
16	29.8	9.9	2034	19	V44806
17	29.8	9.9	2040	21	Z48769
18	29.8	9.9	2045	19	V50436
19	29.8	9.9	2064	21	Z99539
20	29.8	9.9	2084	21	Z56987
21	29.8	9.9	2138	19	V24137
22	29.8	9.9	2143	19	V61937
23	29.8	9.9	2143	20	Z39040
24	29.8	9.9	2188	19	V31375
25	29.8	9.9	2243	19	V52968
26	29.8	9.9	2352	21	Z47926
27	29.8	9.9	12923	10	N90338
28	29.6	9.9	4042	20	X20669
C 29	29	9.7	2297	19	V52836
30	29	9.7	2297	19	V52836
31	29	9.7	2297	20	X86181
32	29	9.7	4428	19	V52837
33	29	9.7	4428	19	V15700
34	29	9.7	4428	20	X86183
35	28.8	9.6	1716	20	Z22975
36	28.8	9.6	13425	19	V52284
37	28.6	9.5	13188	19	V52203
C 38	28.6	9.5	1664976	19	V21209
39	28.4	9.5	3872	19	V49601
40	28.2	9.4	10726	19	V52199
41	28	9.3	1932	21	A26946
C 42	28	9.3	15218	19	V18275
C 43	28	9.3	15218	20	Z22909
C 44	28	9.3	15218	20	Z22914
C 45	28	9.3	15218	20	X35267

ALIGNMENTS

RESULT 1	
X20527/c	
ID X20527 standard; DNA; 10820 BP.	
XX X20527;	
XX 05-MAY-1999 (first entry)	
XX Polynucleotide sequence from the genome of Treponema pallidum.	
XX Treponema pallidum infection; syphilis; Borrelia infection; animal; enzyme production; GS.	
XX Treponema pallidum.	
XX WO9859034-A2.	
XX 30-DEC-1998.	
XX 23-JUN-1998; 98WO-US13041.	
XX 24-JUN-1997; 97US-0050667.	
XX (HUMA-) HUMAN GENOME SCI INC.	
XX Fraser CM;	
XX WPI; 1999-081273/07.	
XX New isolated Treponema pallidum nucleic acids - used to develop products for the detection, diagnosis, characterisation, prevention and therapy of T. pallidum infections, particularly syphilis	
XX Claim 1; Page 343-349; 1150pp; English.	

Human insulin-resp
Human FLAME-1 cDNA
Human MACH related
I-FLICE-1 coding s
Human CLARP coding
Human Casper gene.
Human insulin-resp
Usurpin-alpha poly
Homo sapiens B258
Human CFLIP-L DNA.
Human FLICE-like i
Human FIN-1 encodi
Human G1 protein i
Human apoptosis as
Sequence of human
Polynucleotide seq
Human iduronate-2-
Human IDS cDNA. H
cDNA encoding idur
Human iduronate-2-
Human IDS gene. H
DNA encoding iduro
Brassica napus D22
Streptococcus pneu
Streptococcus pneu
Methanococcus jann
Candida albicans O
Streptococcus pneu
Essential Staphylo
RSV isolate 2B wil
Nucleotide sequenc
Nucleotide sequenc
DNA encoding the L

CC X20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 XX
 SQ Sequence 10820 BP; 2418 A; 2982 C; 2962 G; 2450 T; 8 other;

Query Match 12.1%; Score 36.2; DB 20; Length 10820;
 Best Local Similarity 51.6%; Pred. No. 0.068;
 Matches 83; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 63 ctggcaactactctgttaattgtcattcttaagagcggagggcacacacctgaaccgcgaat 122
 Db 2718 CTTGAGACCCAGCTTTTATTATTGTTATCAAGAGTGGTAGCGCTTGAACACTCAGT 2659

QY 123 ggtctactagaagtacagaagccttcagagatcgggcgctgaattctcgaacacaggg 182
 Db 2658 AATGAGCTTTTCGTGGCACAGCTCTTCGTCAAGCAGCTCTAGAACCGCACACACAGTTC 2599

QY 183 gtgcaattactcaagaaattctctgttgataacactgc 223
 Db 2598 GTGGCAGTTACCGAGCAGCAGCTCCACTTGCAATAATATCC 2558

RESULT 2
 ID V75242/c
 XX V75242 standard; DNA; 543 BP.
 AC V75242;
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE Staphylococcus aureus contig SEQ ID #931.
 KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 361..420
 FT /tag= a
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 XX
 PN EF786519-A2.
 XX
 PD 30-JUL-1997.
 XX
 PF 07-JAN-1997; 97BP-0100117.
 XX
 PR 05-JAN-1996; 96US-0009861.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 XX
 XX WPI; 1997-374922/35.
 DR
 XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
 XX stored on computer readable medium and used in the production of
 PT anti-S. aureus vaccines
 XX
 XX Claim 1; Page 1752-1753; 3271pp; English.
 PS
 XX

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S. aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S. aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S. aureus in a sample. S. aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S. aureus DNA sequences contained on the
 CC computer readable medium.
 XX
 SQ Sequence 543 BP; 178 A; 100 C; 53 G; 152 T; 60 other;

Query Match 10.8%; Score 32.4; DB 18; Length 543;
 Best Local Similarity 50.6%; Pred. No. 0.37; Mismatches 0; Gaps 0;
 Matches 78; Conservative 0;

QY 41 aattgctcaactagggacctggaactggaactctctgttaattgtcttaagagcgg 100
 Db 207 AATTGTACTTGTGCACTAGGACTAACAATCTGTAGTAGCAATCATTTATTGGCCAG 148

QY 101 aggcacacctgaaacccgcgaatggctctactagaagtacagaagccttcagagatgcggg 160
 Db 147 AGAATTGCGTAACCTGGTTACGTTTACTACAAATTTGAACAAGGATTGCTAAGTGCACG 88

QY 161 gctgcaattctcgaacacaggggtgttgcaattact 194
 Db 87 TGTCATATTAGGTAAATAATTAACACAGCAGTTACT 54

RESULT 3
 ID X13336 standard; DNA; 32768 BP.
 XX X13336;
 XX
 DT 19-MAR-1999 (first entry)
 XX
 DE Enterococcus faecalis genome contig SEQ ID NO:399.
 XX
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO9805055-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08985.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Barash SC, Dillon PJ, Kunsch CA;
 XX WPI; 1999-045171/04.
 DR
 XX New isolated Enterococcus faecalis polynucleotides and polypeptides
 XX - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT

PT infection.

Claim 1; Page 1592-1609; 2084pp; English.

A computer readable medium has been developed which has recorded on it 982 nucleotide sequences isolated from the Enterococcus faecalis genome. X12938 to X13919 represent these nucleotide sequences which are primary nucleotide sequences, also known as contigs. The computer-based system can identify fragments of the Enterococcus faecalis genome with commercial importance. The products can be used to detect the presence of Enterococcus faecalis in samples. They can also be used for diagnosing Enterococcus infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of Enterococcus faecalis, or another related organism, in vivo or in vitro. In particular the polypeptides encoded by the Enterococcus faecalis nucleotide sequences can be used in vaccines to prevent or attenuate an Enterococcal infection.

Sequence 32768 BP; 10797 A; 5960 C; 7083 G; 8921 T; 7 other;

Query Match
Best Local Similarity 10.5%; Score 31.6; DB 20; Length 32768;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 52 taggacctgaactggcgaactactctgttaattgtcatttcttaagagcgagcacacctg 111
DB 4846 ttggaccgcccccaaggtaagttcattatggtgttttgaagatacaagcacaccta 4905
QY 112 aaaccgcgaatggtctactagaagtacagaaagccttcagagatgcgggctgcaattct 171
DB 4906 aattcatcaattcatatgaattgtttttgtagatgggaataaacgacagacaaatca 4965
QY 172 cgaacacaggggtgtgcaattac 193
DB 4966 aggtcgtggtgttccatctcc 4987

RESULT 4

X87836
ID X87836 standard; DNA; 579 BP.

AC X87836;

DT 09-NOV-1999 (first entry)

DE S. aureus phosphatidylglycerophosphate synthase pgsA gene.

KW Phosphatidylglycerophosphate synthase; pgsA gene; infection;
KW diagnosis; therapy; vaccine; antibacterial; antibiotic; screening;
KW ss.

OS Staphylococcus aureus.

PN W09942562-A1.

PD 26-AUG-1999.

PF 12-FEB-1999; 99WO-US03092.

PR 20-FEB-1998; 98US-0026017.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PI Holmes DJ, Petit CM, Traini CM, Warren RL;

DR WPI; 1999-518603/43.

DR P-PSDB; Y31662.

XX New synthase polypeptide useful to treat conditions requiring
PT synthase and to screen for antibacterial compounds
XX

PS

Claim 6; Page 42; 47pp; English.

XX The present sequence represents a claimed polynucleotide coding for
CC the phosphatidylglycerophosphate synthase (pgsA, see Y31662) of
CC Staphylococcus aureus WCUH29 (NCIMB 40771). The gene can be
CC isolated from chromosomal DNA of WCUH29 using a hybridisation
CC probe. The invention provides pgsA polynucleotides and
CC polypeptides, vectors, host cells, antibodies and antagonists.
CC The polypeptides can be administered therapeutically, especially by
CC expressing encoding polynucleotides, to treat an individual in need
CC of pgsA polypeptide (claimed). Polynucleotides encoding pgsA are
CC useful for recombinant production of pgsA polypeptides, and can
CC also be used to produce probes for detecting or isolating sequences
CC encoding pgsA or similar sequences e.g. to diagnose diseases
CC relating to polypeptide expression or activity (claimed) and to
CC detect mutations, stage of infection etc. They may also be used
CC for gene mapping. Computer-readable forms of the present sequence
CC may be used to perform homology identification, by comparing the
CC sequence with at least one polynucleotide/polypeptide sequence to
CC identify homology (claimed), or in polynucleotide assembly, by
CC screening for at least one overlapping region between the present
CC sequence and a second polynucleotide sequence (claimed). The
CC polynucleotides can be used as vaccines to induce an immunological
CC response; a mammal is inoculated with a nucleic acid vector
CC directing in vivo expression of the pgsA polypeptide or
CC fragment/variant, in order to produce antibody and/or T cell immune
CC responses to protect the animal from disease (claimed).

XX Sequence 579 BP; 176 A; 62 C; 115 G; 226 T; 0 other;

Query Match
Best Local Similarity 10.3%; Score 30.8; DB 20; Length 579;

Matches 77; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 41 aattgtcactaggacactgaactggcaactactctgttaattgtcatttcttaagagcg 100
DB 270 aattgttactgtgcaactaggactaacaattctgtagtagcaatcattatttgccag 329
QY 101 aggcacacacctgaaacccgcaatggtctactagaagtacagaaagccttcagagatgcgg 160
DB 330 agaatttgcgttaactggtttactgtttactacaaattgcaacaggattgtgaatgcagc 389
QY 161 gctgcaattctcgaacacaggggtgtgtgcaattact 194
DB 390 tggctcaattaggtaaaattaaacagcaggttact 423

RESULT 5

X13479
ID X13479 standard; DNA; 1004 BP.

AC X13479;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:542.

KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX


```
PR 31-OCT-1996; 96US-0029960.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
XX Kunsch CA, Rosen CA;
XX WPI; 1998-272225/24.
XX Computer-readable medium with recorded Streptococcus pneumoniae
XX polynucleotide sequences - useful in diagnostic kits and assays, and
XX pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
XX Claim 1; Page 332-339; 1409pp; English.
XX The present invention describes a computer readable medium which has
XX the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
XX on it, or a representative fragment or a sequence at least 95% identical
XX to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
XX to 391 (V52134 to V52524) are genomic fragments from Streptococcus
XX pneumoniae. The present invention also describes an isolated nucleic acid
XX molecule encoding a homologue of any of the fragments of the S.pneumoniae
XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
XX by a process comprising: (a) screening a genomic DNA library using as a
XX probe a target sequence defined by any of the sequences in SEQ ID NO:1
XX to 391, identifying members of the library which contain sequences
XX that hybridize to the target sequence and isolating the nucleic acid
XX molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
XX from an organism, amplifying nucleic acid molecules whose nucleotide
XX sequence is homologous to amplification primers derived from the
XX fragment of the S. pneumoniae genome to prime the amplification and
XX isolating the amplified sequences. The computer readable medium can be
XX used in a computer-based system for identifying fragments of the
XX S. pneumoniae genome of commercial importance, or expression modulating
XX fragments of the S. pneumoniae genome. Products from the present
XX invention can be used in diagnosis kits and assays, and pharmaceutical
XX compositions and vaccines for S. pneumoniae.
XX Sequence 13206 BP; 4016 A; 2929 C; 2473 G; 3786 T; 2 other;

Query Match 10.28; Score 30.6; DB 19; Length 13206;
Best Local Similarity 53.8%; Pred. No. 4.8;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 12 aacaccgatacctgctgggattgatacacaattgctcaactagtaggacctggaactggaact 71
   || || || || || || || || || || || || || || || || || || || || ||
Db 9898 aatccactccctctcggcatgacacagggttaattcttcattgcttggcgaagtgtat 9957

Qy 72 actcttgtaattgtcattcttaagagcgaggacacacctgaaacccgcaatggtcta 128
   | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9958 aatctgtagtgcctttttgagcagcgacatcattcatacctaaagggtccta 10014

RESULT 8
W74475
ID V74475 standard; DNA; 22243 BP.
XX
XX V74475;
XX
XX 16-MAR-1999 (first entry)
XX
XX Staphylococcus aureus contig SEQ ID #164.
XX
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
XX
XX Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH 21661..21720
```

```
FT misc_feature 61..120
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 1861..1920
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 3661..3720
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 5461..5520
FT /*tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 7261..7320
FT /*tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 9061..9120
FT /*tag= f
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 10861..10920
FT /*tag= g
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 12661..12720
FT /*tag= h
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 14461..14520
FT /*tag= i
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 16261..16320
FT /*tag= j
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 18061..18120
FT /*tag= k
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 19861..19920
FT /*tag= l
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT
FT misc_feature 21661..21720
```

FT /*tag= m
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP786519-A2.
XX
XX
XX 30-JUL-1997.
XX
XX 07-JAN-1997; 97EP-0100117.
XX
XX 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
XX Claim 1; Page 800-813; 327lpp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX
XX Sequence 22243 BP; 7693 A; 2910 C; 4245 G; 5604 T; 791 other;
SQ

Query Match 10.2%; Score 30.6; DB 18; Length 22243;
Best Local Similarity 50.7%; Pred. No. 5.9;
Matches 72; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 2 attattgacaacacagatcctcgtcgggattgatacatcaaatgtcacaaggacctga 61
Db 20692 atttttgaataacggctctcgtcgaatgaaatgattatcgaatacaataaata 20751
QY 62 actggaacactcttgtaattgtcattcttaagagcgaggacacacacacccgcaa 121
Db 20752 gctgttacgctcgtcgcacatattattccaaatacaaaaagattcattcgaagtca 20811
QY 122 tggcttactagaagtacagaaa 143
Db 20812 tggcttactagaagtacagaaa 20833
RESULT 9
ID Z39041
XX Z39041 standard; cDNA; 1062 BP.
AC Z39041;
XX Z39041;
DT 25-FEB-2000 (first entry)

XX Human FLICE-like inhibitory protein short form nucleotide sequence.
DE
XX
KW Human; FLICE-like inhibitory protein long form; FLIP-L; FLIP-S;
KW FLICE-like inhibitory protein short form; apoptosis inhibitor;
KW arteriosclerosis; vascular wall inflammation; vascular injury;
KW Fas ligand-mediated apoptosis; atherosclerosis; transplant; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 294..959
FT /*tag= a
FT /product= "FLICE-like inhibitory protein short form"
FT /note= "FLIP-S; apoptosis inhibitor"
XX
XX WO9942570-A1.
XX
XX 26-AUG-1999.
XX
XX 19-FEB-1999; 99WO-US03558.
XX
XX 20-FEB-1998; 98US-0075471.
XX
XX (SELI-) ST ELIZABETH'S MEDICAL CENT BOSTON INC.
XX
XX Walsh K;
XX
XX WPI; 1999-527469/44.
XX P-PSDB; Y57455.
XX
XX Treating conditions characterized by vascular wall inflammation
XX
XX Example 2; Page 71; 105pp; English.
XX
XX The present sequence encodes human FLICE-like inhibitory protein short
CC form, designated FLIP-S. The present invention describes a new treatment
CC of a condition characterised by vascular wall inflammation in a subject
CC comprising administering a FLIP molecule to inhibit Fas ligand-mediated
CC apoptosis of vascular endothelial cells in the subject. The method can
CC be used to treat atherosclerosis, transplant arteriosclerosis and
CC vascular injury.
XX
XX Sequence 1062 BP; 328 A; 212 C; 246 G; 276 T; 0 other;
SQ

Query Match 9.9%; Score 29.8; DB 20; Length 1062;
Best Local Similarity 51.1%; Pred. No. 3.3;
Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 54 ggacctgaactggaactactcttgaattgtcatttctaagagcgaggacacacctgaa 113
Db 246 ggagctgtactgcaagacccttgtgagcttcccttagtctaagagtagtagtctgctgaa 305
QY 114 accgcaatggtctactagaagtacagaaagccttcagagatcggggctgcaattctcg 173
Db 306 gtcatccatcagggttggaagcacttgatcacagatgagaagagagatgctgctcttttg 365
QY 174 aacacaggggtgtgcaat 190
Db 366 tgcgggagattgtctat 382
RESULT 10
ID V61936
XX V61936 standard; DNA; 1190 BP.
AC V61936;
XX
XX 12-JUL-1999 (first entry)
XX Human CFLIP-S DNA.
XX

CC overlapping EST clones were sequenced to deduce the sequence of
CC the full-length MRIT alpha 1 clone and 2 isoforms, alpha 2 (see
CC X25509) and beta 1 (see X25510). The MRIT gene was localised to
CC the human chromosome 2q32-33 region. MRIT alpha 1 includes an
CC N-terminal death effector domain and a C-terminal caspase homology
CC domain, but is not a cysteine protease. Selective enhancers and
CC inhibitors of MRIT apoptotic activity can be identified and used
CC to treat diseases mediated by the dysfunction of programmed cell
CC death or proliferation. In addition, nucleic acid molecules
CC encoding pro-apoptotic isoforms of MRIT can also be used as
CC inducers of programmed cell death. A cell accumulation disorder
CC such as cancer, autoimmune disease, viral infection, angiogenesis
CC and atherosclerosis is treated by administering an agent that
CC selectively enhances MRIT apoptotic activity, thereby inducing
CC apoptosis in a subject. A disorder of cell loss, such as a
CC neurodegenerative disorder, including Alzheimer's disease,
CC Parkinson's disease, retinitis pigmentosa, stroke, aplastic
CC anaemia, myocardial infarction or AIDS can be treated by
CC administering an agent that selectively inhibits MRIT apoptotic
CC activity.
XX
SQ Sequence 1996 BP; 543 A; 461 C; 540 G; 451 T; 1 other;

Query Match 9.9%; Score 29.8; DB 20; Length 1996;
Best Local Similarity 51.1%; Pred. No. 4.2;
Matches 70; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Qy 54 ggacctgaactggcaactactctgttaattgtcattcttaagagcgaggacacactgaa 113
Db ||| ||| ||| | || || | ||||| | | ||||
387 ggagctgtactgcaagacccttgagctccctagctcctaagtaggtgtctgtgaa 446

Qy 114 accgcaatggtctactagaagtacagaagccttcagagatcgggggtgcaattctcg 173
Db | | | | | ||||| || | | | | ||||| |||
447 gtcatccatcagggttgaaagacattgtatcacagatgagaaggagatgctgtcttttg 506

Qy 174 aaacagggtgttgcaat 190
Db | | | ||||| ||
507 tgccgggatgtgtctat 523

Search completed: November 4, 2000, 13:45:55
Job time: 16482 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:34:22 ; Search time 189.35 Seconds
(without alignments)
239.622 Million cell updates/sec

Title: US-09-300-482-619
Perfect score: 300
Sequence: 1 gattattgacacacgat.....ttcagaaatgtctgtgtgg 300

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5E_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/6_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29.8	9.9	1750	5	US-08-859-167-1
2	29.8	9.9	1750	5	US-09-109-273-1
3	29	9.7	2297	2	US-08-484-493-1
4	29	9.7	2297	2	US-08-484-493-1
5	29	9.7	2297	2	US-08-484-493-1
6	29	9.7	2297	2	US-08-484-493-1
7	29	9.7	4428	2	US-08-484-493-6
8	29	9.7	4428	2	US-08-484-493-6
9	28	9.3	1932	5	US-08-345-212-6
10	28	9.3	1932	5	US-08-714-918-97
11	27.6	9.2	1146	1	US-08-892-403A-2
12	27.6	9.2	1146	1	US-08-487-810-1
13	27.6	9.2	3906	3	US-08-469-537A-102
14	27.6	9.2	4165	2	US-08-442-248-1
15	27.4	9.1	1055	2	US-08-440-815-1
16	27.4	9.1	1055	2	US-08-618-464-1
17	26.8	8.9	1055	6	PCT-US95-04896-1
18	26.8	8.9	2441	1	US-08-332-312-1
19	26.8	8.9	2624	1	US-08-032-382B-1
20	26.8	8.9	3705	6	PCT-US96-03940-7
21	26.4	8.8	5648	6	PCT-US96-03940-1
22	26.4	8.8	1806	4	US-08-980-060-12
23	26.4	8.8	3089	1	US-08-980-060-7
24	26.4	8.8	3459	4	US-07-921-796-7
25	26.4	8.8	6030	2	US-08-980-060-3
26	26.4	8.8	8700	3	US-08-392-625-16

ALIGNMENTS

RESULT 1
US-08-859-167-1
; Sequence 1, Application US/08859167
; Patent No. 6037461
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: FADD-LIKE ANTI-APOPTOTIC MOLECULES, METHODS OF
; TITLE OF INVENTION: USING THE SAME, AND COMPOSITIONS FOR AND METHODS
; TITLE OF INVENTION: OF MAKING THE SAME
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 6037461ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/859,167
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1750 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 413..1750
US-08-859-167-1

Sequence 16, Appl
Sequence 18, Appl
Sequence 1, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 5, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 35, Appl
Sequence 27, Appl
Sequence 1, Appl
Sequence 5, Appl
Sequence 25, Appl
Sequence 8, Appl
Sequence 11, Appl
Sequence 1, Appl

Query Match 9.9%; Score 29.8; DB 5; Length 1750;
Best Local Similarity 51.1%; Pred. No. 0.85;

APPLICANT: Wilson, Peter J

APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
IDURONATE 2-SULFATASE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/484,493
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 84162
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 4428 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 332..434
FEATURE:
NAME/KEY: intron
LOCATION: 536..537
FEATURE:
NAME/KEY: CDS
LOCATION: 693..829
FEATURE:
NAME/KEY: intron
LOCATION: 962..963
FEATURE:
NAME/KEY: CDS
LOCATION: 1044..1221
FEATURE:
NAME/KEY: intron
LOCATION: 1350..1351
FEATURE:
NAME/KEY: CDS
LOCATION: 1480..1569
FEATURE:
NAME/KEY: intron
LOCATION: 1716..1717
FEATURE:
NAME/KEY: CDS
LOCATION: 1841..2041
FEATURE:
NAME/KEY: intron
LOCATION: 2206..2207

FEATURE:
NAME/KEY: CDS
LOCATION: 2294..2464
FEATURE:
NAME/KEY: intron
LOCATION: 2585..2586
FEATURE:
NAME/KEY: CDS
LOCATION: 2684..2810
FEATURE:
NAME/KEY: intron
LOCATION: 2904..2905
FEATURE:
NAME/KEY: CDS
LOCATION: 3033..3206
FEATURE:
NAME/KEY: intron
LOCATION: 3308..3309
FEATURE:
NAME/KEY: CDS
LOCATION: 3435..3908
US-08-484-493-6

Query Match 9.7%; Score 29; DB 2; Length 4428;

Best Local Similarity 53.0%; Pred. No. 2.5;

Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 121 atggtctactagagtcacagaaagcccttcagagatgcggggctgcaattctcgaacagg 180

Db 3636 ATTGCCTATAGCCAGTATCCCGCGCTTCAGACATCCCTCAGTGGAAATTCGACAGCG 3695

QY 181 gtgtgcaattactcaagaaattctctgttgatacaactgctagaatagaggat 237

Db 3696 AGTTTAAAGATATAAAGATCATGGGCTATTCCATAGCACCACCATAGTAT 3752

RESULT 7

US-08-484-494-6
Sequence 6, Application US/08484494

Patent No. 5798239

GENERAL INFORMATION:

APPLICANT: Wilson, Peter J

APPLICANT: Morris, Charles P

APPLICANT: Anson, Donald S

APPLICANT: Occhiodoro, Teresa

APPLICANT: Bielicki, Julie

APPLICANT: Clements, Peter R

APPLICANT: Hopwood, John J

TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF

IDURONATE 2-SULFATASE

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Scully, Scott, Murphy & Presser

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: New York

COUNTRY: USA

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,494

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 991,973

FILING DATE: 17-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Digiglio, Frank S

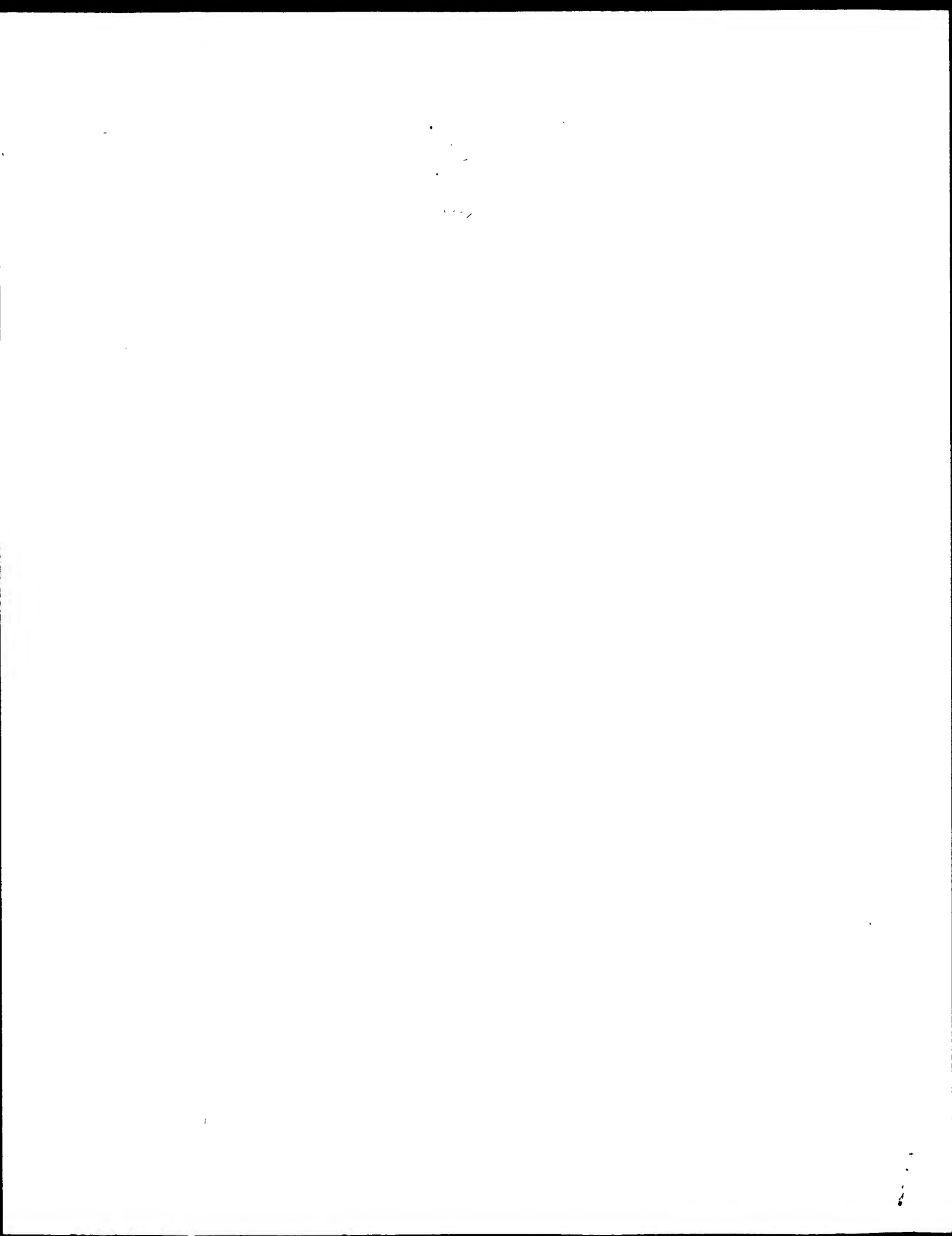

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1055 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (clonal)
; IMMEDIATE SOURCE:
; CLONE: E. coli K12 dhaj polynucleotide (cdna)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1054
; US-08-618-464-1
```

```
Query Match          9.1%; Score 27.4; DB 2; Length 1055;
Best Local Similarity 62.3%; Pred. NO. 4.4;
Matches 43; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 150 agagatgcggggctgcaattctcgaaacagggtgttgcaattactcaagaaaaattctctg 209
    || || | | | | | | | | | | | | | | | | | | | |
Db 740 AGCGAAGTTGATCGGGACTTCGCAATACAGGTTGCTTCACGCTCGAAAAATCGGGTG 681

QY 210 ttggataac 218
    || || | | |
Db 680 CTGTTTAAAC 672
```

```
Search completed: November 4, 2000, 13:34:28
Job time: 16704 sec
```



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:37:55 ; Search time 4075.18 Seconds
(without alignments)
412.624 Million cell updates/sec

Title: us-09-300-482-356

Perfect score: 385

Sequence: 1 caaccggcaccagcctgac.....acggtctgggagttttttt 385

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_bal.*

2: gb_ba2.*

3: gb_om.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pil.*

8: gb_pl2.*

9: gb_pr1.*

10: gb_pr2.*

11: gb_pr3.*

12: em_fun.*

13: em_hum1.*

14: em_hum2.*

15: em_in.*

16: em_om.*

17: em_or.*

18: em_ov.*

19: em_pat.*

20: em_ph.*

21: em_pl.*

22: em_ro.*

23: em_sts.*

24: em_sy.*

25: em_un.*

26: em_vil.*

27: gb_htg1.*

28: gb_htg2.*

29: gb_in1.*

30: gb_in2.*

31: em_bal.*

32: em_ba2.*

33: em_hum3.*

34: em_hum4.*

35: gb_pr4.*

36: gb_htg3.*

37: gb_htg4.*

38: gb_htg5.*

39: gb_htg6.*

40: gb_htg7.*

41: em_htg1.*

42: em_htg2.*

43: em_htg3.*

44: em_hum5.*
45: gb_pl3.*
46: gb_pr5.*
47: gb_htg8.*
48: gb_htg9.*
49: gb_htg10.*
50: gb_htg11.*
51: gb_htg12.*
52: gb_htg13.*
53: gb_htg14.*
54: gb_in3.*
55: gb_htg15.*
56: gb_htg16.*
57: gb_htg17.*
58: em_htg4.*
59: em_htg5.*
60: em_htg6.*
61: em_htg7.*
62: em_hum6.*
63: gb_htg18.*
64: gb_htg19.*
65: gb_ba3.*
66: em_htg8.*
67: em_htg9.*
68: em_htg10.*
69: em_htg11.*
70: em_htg12.*
71: em_htg13.*
72: em_htg14.*
73: em_htg15.*
74: em_htg16.*
75: em_htg17.*
76: em_htg18.*
77: em_htg19.*
78: em_htg20.*
79: em_htg21.*
80: em_htg22.*
81: em_htg23.*
82: gb_pr6.*
83: gb_pr7.*
84: gb_htg20.*
85: gb_htg21.*
86: gb_htg22.*
87: gb_htg23.*
88: gb_ro.*
89: gb_sts1.*
90: gb_sts2.*
91: gb_sy.*
92: gb_un.*
93: gb_vil.*
94: gb_vil2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	247.8	64.4	101441	8 AP002542	AP002542 Oryza sat
2	187.8	48.8	2319	8 CAY15781	Y15781 Capsicum an
3	177.8	46.2	1802	8 CPTKT3	Z46646 C.plantagin
4	170.4	44.3	2629	5 A52295	A52295 Sequence 1
5	170.4	44.3	2629	5 AR072009	AR072009 Sequence
6	170.4	44.3	2629	5 E13696	E13696 cDNA encodi
7	167	43.4	2447	45 STTKTMR	Z50099 S.tuberosum
8	165.4	43.0	2232	8 CPTKT7	Z46647 C.plantagin
9	163.8	42.5	2207	8 CPTKT10	Z46647 C.plantagin
10	162.4	42.2	2519	45 SPITRAN	L76554 Spinacia ol
11	162.2	42.1	107865	8 ATT4C21	AL162295 Arabidops
12	155.6	40.4	2595	8 CPA249787	AJ249787 Cyanophor

[illegible]

ORGANISM	REFERENCE
Craterostigma plantagineum	
Eukaryota; Viridiplantae	
Streptophyta; Embryophyta; Tracheophyta;	
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;	
Asteridae; Gentianaceae; Lamiales; Scrophulariaceae; Craterostigma.	
1 (bases 1 to 2232)	

REFERENCE	1 (bases 1 to 2232)
AUTHORS	Bartels, D.
TITLE	Direct Submission
JOURNAL	Submitted (10-NOV-1994) Bartels D., Max-Planck-Institut fuer Zuerchtforschung, Pflanzenzuechtung und Ertragsphysiologie, Carl-von-Linne-Weg 10, Cologne, NRW, Germany, 50829
REFERENCE	2 (bases 1 to 2232)
AUTHORS	Bernardini, G., Schwall, G., Lottspeich, F., Salami, F. and

REFERENCE	(to be added by the author)
AUTHORS	Bernacchia, G., Schwall, G., Lottspeich, F., Salamini, F. and Bartels, D.
TITLE	The transketolase gene family of the resurrection plant <i>Cratogeomys plantaginum</i> : differential expression during the rehydration phase
JOURNAL	EMBO J. 14 (3), 610-618 (1995)
MEDLINE	95163594

```

MEDLINE      95163594
FEATURES
  source      Location/Qualifiers
    1..2232
      /organism="Craterostigma plantagineum"
      /db_xref="taxon:4153"
      /clone="Cpl-1kt7"
    29..2059
      /gene="Cpl-1kt7"
    29..2059
      /gene="Cpl-1kt7"
      /EC_number="2.2.1.1"
      /citation={2}
      /codon_start=1
      /product="transketolase"
      /protein_id="CAA86609.1"
      /db_xref="GI:664903"

```

```

525 a 586 c 680 q 441 t
BASE COUNT
/translation="MAPIKTIAPSLVSKSVNTRIFLADAVEKASKGHPGMPMGCG
/ob_xref="GI:664903"
PQWGSYKDFNRPKNPWNFRNDRFLVSNAGHGMQLQYALHLISGYDSVKEEDLKH
MGHSTPAHPENFTPGVEVTPGLGQIASAVGLAAVEKHLAARYNKPGEFVDSH
VYLGDCQMEGVSNCAASHAHGLGLIALYDDNHLITDGTDFVATDEBDVDRYGR
LGWHVYVYKNGDCCDEIRALBEAKSVKDRPTMKVTTTIGYGAPSKANTYGVHG
LGPKEAETRYVDPDYFPPHVDYKQWHSRHIAEGAALESANNAKFAEQKFFPB
AADLSITGSELPTNWESIPTYTPNGLPRTLHQIILGDLVLPGLUGSADSI
LSNFAKNSGDFQKSGPERNVPFGAREHAMGSCINGLALHSPGLYPCATYFVF
VYRAAMRTSALS KARVLYIMTHDSIGLEDGTPHOYEHLASFRAMENILTLRPADQ
ETAGAYRAAVQNGERPSTLYLAROKLPOLPGTSTEGVSKGYVILSDNSRGNKSPDV
LIGTASLELAEIRALOKPQKGVVVVSLVCWELPABQSEKYRETIVLPQSVTARV
EAGSTGWERFTGPKQKAVGIDTRGASAPARLKFRTGIVTAEVAVAAAKEIC"

```

BASE COUNT	525 a	586 c	680 g	441 t	ORIGIN
Query Match					
Best Local Similarity					
Matches 232;					
Conservative 0;					
Mismatches 111;					
Indels 0;					
Gaps 0;					
Score 165.4;					
DB 8;					
Length 2232;					

	Matches	232;	Conservative	0;	Mismatches	111;	Indels	0;	Gaps
Qy	12	aagcctgacatcatgtggttggccaccgcctcgagctgagatcgagggaatacggcc	71						
Db	1712	AAGCCGACGTCACTACTGATCGGGACGGCTCGGAGCTTGAGATAGCGGTAGGCCCGGG	1771						
Qy	72	gaagagctgaggaaaggagggggaagaagcgttcgcgttgtctcttctctctcttggaactc	131						

Qy 72 **gacgagctgaggaagagaggaagcgtccgcgctgctcttccgtctccctcggaac** 131
Db 1772 **GACGAGCTCAGGAAGAGGGCAAAAAGGTC**CGAGTGTCTCGCTTGTGTGTGGAGATTG 1831
Qy 132 **tttgatgacagctcgatcagtcacaaagagagcgtcctccctccgcgacgtcacagcgag** 191
Db 1832 **TTCCCGCAGCAGTCCGAGAGTATACAGAGACGGTTT**GCCCTCGGGGTTTACGCGTAGG 1891

Db	1832	TTGCGGAGCAGTCCGAGAAAGTATAGAGAGACGGTTTCCTCGGGGGTATAGGCGCAGG	1891
Qy	192	atcagcatcgagccgggtgccactctcggctggcagaagtacgtgcagcccaaggaag	251
Db	1892	GTTAGCGCTGGAGCGGGCTCGACTTTCGGGTGGGAGAGGTTTCATCGGGCCGAAGGTAAG	1951
Qy	252	gccattggcatcgacaaagttcgcgcgagtgctcctgcgcggagcacgatctcacaaaggagtac	311

[illegible]

Qy		312 ggcattcacccgtgagagcatcattgcactgccaaaggacgtttt	354
Db		2012 GGGATAACCGTCGAAGCACTGTCGGCGGCAGGAGATT	2054

Db	2012	GGGATACACCGTCGAAGCAGTCGTCGCGGCGCGAAGGAGATTT	2015
RESULT	9		
CPTKT10			
LOCUS	CPTKT10	2207 bp	mRNA
DEFINITION	C.plantagineum tkt10 gene for transketolase.		

DEFINITION	C.plantagineum tkt10 gene for transketolase.
ACCESSION	Z46647
VERSION	246647.1 GI:664898
KEYWORDS	tkt10 gene; transketolase.
SOURCE	Craterostigma plantagineum.
ORGANISM	Craterostigma plantagineum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Charophyta; Lemnaceae; Lemnaceae

REFERENCE
BARTELS, D.
TITLE
JOURNAL
Submitted (10-NOV-1994) Bartels D., Max-Planck
Zentrum für Tierphysiologie, Universität Erlangen
und Nürnberg, 91054 Erlangen, Germany

JOURNAL
Submitted (10-Nov-1994) Bartsch D., Max-Planck
Züchtungsforschung, Pflanzenzüchtung und Ernt-
Carl-von-Linne-Weg 10, Cologne, NRW, Germany,
2 (bases 1 to 2207)
Barnackia, G., Schwall, G., Lottspeich, F., Salas
Bartsch, D.
AUTHORS
TITLE
The transketolase gene family of the resurrected
Craterostigma plantaginum: differential expres-

JOURNAL	craterostigma plantagineum: differential exper
MEDLINE	EMBO J. 14 (3), 610-618 (1995)
FEATURES	rehydration phase
	95163594
	Location/Qualifiers
source	1. .2207
	/organism="Craterostigma plantagineum"

```

/organism="Craterostigma plantagineum"
/db_xref="taxon:4153"
/clone="Cpl-1kt10"
3. .2042
/gene="Cpl-1kt10"
3. .2042
/gene="Cpl-1kt10"
/EC_number="2.2.1.1"
/citation={2}
/codon_start=1
/product="transketolase"
/protein_id="CAA86608.1"
/db_xref="GI:604899"

```

```

/db_xref="GI:664899"
/db_xref="SWISS-PROT:Q42675"
/translation="MAKTPSPSSAAAAAEILVKVSVN
MCPAMGHVLFDERMKNPKYFNFRDVLGAGH
KGLRWGSKTPAHPEVFTPCVETGTPLGQGYLA
DKHYTYMILGDCGCMGIESNASSLAAHWGLGLA
RREALCWHVLYVANGDGYDEIRAIKVAKSVTDKPK

```

REALGWSHVVANGNDGJDEIRGAIKVAVSVTDUKR
HGNALGPDEAETRNQLNGCYETHFVDDVKKHWSR
YFKEAAELKSTTIGELPLGWEKALPTYPESPNGPNT
ADITANMAFLSGDSDFOKETPTGRNLKFGAREHGM
VFDYMRRAAIVTAALSKARVVIMTHDSIGLGDGP
ADGNETAGAYKVAENAGRPSTLSLSPKLPOLPGDGT
EVLMTGTSSELEFIIARAGEKDEKCKKVVSVSWV
HYVBAGSGEPPFQFICBAWGTDBFCASASAPYI

BASE COUNT 520 a 582 c 682 g 423 t

ORIGIN

Query Match 42.5%; Score 163.8; DB 8; Len
Best Local Similarity 68.1%; Pred. No. 2; 6e-21;

Query: f06879
Best local similarity 68.1%; Pred No. 2, 6e-21;
Matches 22; Conservative 0; Mismatches 107; Incon-

QY 12 aaagctgacatcatgttggttggcaacgcgactccgagctggagatcgc
||||| |||| |
DB 1695 AAGCCGAGGTGATCTGATTGGGGACGGCTCGGAGCTTGATAGTACC

Qy 72 gacgagctgaggaaggagggaagacggtccgcgctcgtctcgtt

A-1030 Vienna, AUSTRIA

FEATURES
source 1..2595
 /organism="Cyanophora paradoxa"
 /strain="Pringsheim"
 /db_xref="taxon:2762"
 /clone="tk15"
transit_peptide 25..294
 /gene="tkc"
CDs
 25..2340
 /gene="tkc"
 /EC_number="2.2.1.1"
 /function="pentose phosphate (Calvin) cycle"
 /citation=[1]
 /codon_start=1
 /product="putative transketolase precursor"
 /protein_id="CAN58135.1"
 /db_xref="GI:6048212"
translation="MSAFVVPSPGSSSSAIVASSQVISREHAPKIHRSRAQNFAID
TRMFLPNTPSQAQRFQFNGLRRAESVFSSKSDANAALRQSXVAASTGFVINNOATA
AMPVDKDCINSIRFLAIDAIEAKAGSHGLPMGCAPMSYVLNPEFMKNPKNPMDIWD
RDFVLSAGHGCMLOYALLYLTGYDSVGIEDIKTFQWSESECPGHPENFTKGIEVTT
GPUGQAGAGVLAEGAEHLTAARFNPAVKLDIHTYYIMGDCNMGEVAAEASLAG
HYGLNLIALDYDNEISIGDTISFDTEVTKRYSGYGHMTVVVEDGNTDINAIRKAI
AEAARYDKPLSVTIRTTIGYSPNKANSYAVHGAALGDKEVDATRONLGNPYAPEFI
PEAMNKRWAIEFKGSADENWKFAEKYKAPQEAADFKEYVMKKELPANMKCLP
VYPDNDATGATRIKLSTENLNATADAPTFLGGSADLIASNMTLLKKYADFOKTSPEG
RNLRFGVREFMAAIAINGLHLHSGLIPYGATFLVSDDYMRAAMLSSLSKCRTIYVL
THDSIGAGEDPTQHPIEHUAHAHVNTVFPRPCDNVSGSYGAIKVAVBERETPSLMI
LTRQTILDGTSIANTGAYLVSDNSTNGKPDLILMGTVSGEVHLCVNAAEAIRKEG
KTVRVSMSPWELFERQSAEYEASFVKDC KRALAVEAASFSFGWHKYFGDEGANVSDI
GFCASAPGDKLMQEFGETTENVIQAQKLLA"

gene
 25..2340
 /gene="tkc"
mat_peptide
 295..2337
 /gene="tkc"
 /EC_number="2.2.1.1"
 /product="putative transketolase precursor"

BASE COUNT 501 a 922 c 731 g 441 t
ORIGIN

Query Match
Best Local Similarity 40.4%; Score 155.6; DB 8; Length 2595;
Matches 227; Conservative 0; Mismatches 119; Indels . 0; Gaps 0;

Qy 3 accggaccgaagccgtgacatcaattgggttgggcaccgccgtcccgagtggagatcgcgggc 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1981 ACCAACGCCAAGACCGCACCTGATCCTCATGGCCACTGGCTGGAGAGTGCACTCTCGGTC 2040
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 63 aatcgccgcagagctgaggaaaggggaagacggttcgctcgcttgcttgtcgcttc 122
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2041 AACCGCGCGAGGCGATCCCGCAAGAGGGGCAAGACCCTGGCGCTGTGTGCA/GCCCTCC 2070
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 123 tgggaactctttgatgagcatcgatgagtagtacaaagagagcgtccctccctgccgagctc 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2101 TGGGAGCTTTTCGAGCGCCAGTCGGCTGAGTAGACAGGCGCTGTGCTTCCCCAAGGACTGC 2160
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 183 acagcgagatcacgatacgagccgggtcccactctcgctggcgagaagatcgctcgagacc 242
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2161 AAGAGGCCCTTCGCCGTCGAGGCCGCTCGAGCTTCGGCTGGCAACAAGTACTTCGGGCGAC 2220
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 243 caaggcaaggccattggcagatcacaaagtctcgggcgagtgctctctcgcggaagatctac 302
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2221 GAGGCGCGATGTTCTCCATCGACGGCTTCGGCGCGTCTGGCGCGTCCGGCGGACAAAGCTCATG 2280
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 303 aaggagtaaggcataccgctggagagcatcatcattgcaactgtccaaga 348
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2281 CAGGAGTTCGGCTTCACACCGGAAAGTGTATCCGCCCGGCGCAAGA 2326

RESULT 13
AC002387

exon	1	9346	9347	9348	9349	9350	9351	9352	9353	9354	9355	9356	9357	9358	9359	9360	9361	9362	9363	9364	9365	9366	9367	9368	9369	9370	9371	9372	9373	9374	9375	9376	9377	9378	9379	9380	9381	9382	9383	9384	9385	9386	9387	9388	9389	9390	9391	9392	9393	9394	9395	9396	9397	9398	9399	9400	9401	9402	9403	9404	9405	9406	9407	9408	9409	9410	9411	9412	9413	9414	9415	9416	9417	9418	9419	9420	9421	9422	9423	9424	9425	9426	9427	9428	9429	9430	9431	9432	9433	9434	9435	9436	9437	9438	9439	9440	9441	9442	9443	9444	9445	9446	9447	9448	9449	9450	9451	9452	9453	9454	9455	9456	9457	9458	9459	9460	9461	9462	9463	9464	9465	9466	9467	9468	9469	9470	9471	9472	9473	9474	9475	9476	9477	9478	9479	9480	9481	9482	9483	9484	9485	9486	9487	9488	9489	9490	9491	9492	9493	9494	9495	9496	9497	9498	9499	9500	9501	9502	9503	9504	9505	9506	9507	9508	9509	9510	9511	9512	9513	9514	9515	9516	9517	9518	9519	9520	9521	9522	9523	9524	9525	9526	9527	9528	9529	9530	9531	9532	9533	9534	9535	9536	9537	9538	9539	9540	9541	9542	9543	9544	9545	9546	9547	9548	9549	9550	9551	9552	9553	9554	9555	9556	9557	9558	9559	9560	9561	9562	9563	9564	9565	9566	9567	9568	9569	9570	9571	9572	9573	9574	9575	9576	9577	9578	9579	9580	9581	9582	9583	9584	9585	9586	9587	9588	9589	9590	9591	9592	9593	9594	9595	9596	9597	9598	9599	9600	9601	9602	9603	9604	9605	9606	9607	9608	9609	9610	9611	9612	9613	9614	9615	9616	9617	9618	9619	9620	9621	9622	9623	9624	9625	9626	9627	9628	9629	9630	9631	9632	9633	9634	9635	9636	9637	9638	9639	9640	9641	9642	9643	9644	9645	9646	9647	9648	9649	9650	9651	9652	9653	9654	9655	9656	9657	9658	9659	9660	9661	9662	9663	9664	9665	9666	9667	9668	9669	9670	9671	9672	9673	9674	9675	9676	9677	9678	9679	9680	9681	9682	9683	9684	9685	9686	9687	9688	9689	9690	9691	9692	9693	9694	9695	9696	9697	9698	9699	9700	9701	9702	9703	9704	9705	9706	9707	9708	9709	9710	9711	9712	9713	9714	9715	9716	9717	9718	9719	9720	9721	9722	9723	9724	9725	9726	9727	9728	9729	9730	9731	9732	9733	9734	9735	9736	9737	9738	9739	9740	9741	9742	9743	9744	9745	9746	9747	9748	9749	9750	9751	9752	9753	9754	9755	9756	9757	9758	9759	9760	9761	9762	9763	9764	9765	9766	9767	9768	9769	9770	9771	9772	9773	9774	9775	9776	9777	9778	9779	9780	9781	9782	9783	9784	9785	9786	9787	9788	9789	9790	9791	9792	9793	9794	9795	9796	9797	9798
------	---	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------


```

LOCUS      AC002387      122871 bp      DNA      PLN      05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 242 of 255 of the
complete sequence. Sequence from clones T14P1, F4L23.
ACCESSION  AC002387 AE002093
VERSION     AC002387.2 GI:6598365
KEYWORDS   HFG.
SOURCE      thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsids.
REFERENCE  1 (bases 1 to 122871)
            Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D.,
            Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E.,
            Feldblum, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Ronning, C.M.,
            Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., Vankken, S.E., Umayam, L.,
            Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H.,
            Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D.,
            Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and
            Venter, J.C.
            Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
            Nature 402 (6763), 761-768 (1999)
JOURNAL    Nature 402 (6763), 761-768 (1999)
MEDLINE    20083487
REFERENCE  2 (bases 1 to 122871)
            Lin, X.
            Direct Submission
            Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
            On Dec 17, 1999 this sequence version replaced gi:2583106.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).
COMMENT    Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (http://arthur.epm.org/pub/pub/xgrail), GeneFinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
            (http://www.cds.dtu.dk/services/NetGene2/), searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on those analyses.
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by two
            or more gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/repeatMasker.html). Genes are
            numbered from the top to bottom of the chromosome.

            We thank the GSHU/WashU/ABI consortium for sequencing BAC clones
            F6P23, F5J6, T1A5, and T13L16, the ESSA group for sequencing clones
            F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
            and Satoshi Tabata for helpful assistance. In addition, we would
            like to thank the TIGR Bioinformatics Department, especially Lixin
            Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
            Peterson, Michael Holmes, and Delwood Richardson for software and
            database support.

            This work was supported by the National Science Foundation,
            Department of Energy and the US Department of Agriculture.

            Address all correspondence to: at@tigr.org.
            Location/Qualifiers
            1..122871
               /organism="Arabidopsis thaliana"
               /cultivar="Columbia"

```

```

/db_xref="taxon:3702"
/chromosome="II"
/complement(<1..2952)
/note="Sequence from clone T14P1"
/rpt_family="(TAAAA)n"
/join(2227..2322,2480..2527,2650..2705,3098..3150,
3250..3641)
/gene="At2g45170"
2227..3641
/gene="At2g45170"
/note="T14P1.2"
/join(2488..2527,2650..2705,3098..3150,3250..3469)
/gene="At2g45170"
/codon_start=1
/product="putative microtubule-associated protein"
/protein_id="AAF18602.1"
/db_xref="GI:6598366"
/translation="MNRGSIKFKMDDDFEKKRAEAGRIREKYPDRIPVIVEKAEKSEVP
NIDKKYLVPSDLTVGQFVVVIRKRIKLSAEKALIFVDNVLPTGELMSSVYEDKKD
EDGLYITYSGENTEGASSI"
/complement(2953..>122871)
/note="Sequence from clone F4L23"
4282..4870
/gene="At2g45180"
4282..4870
/gene="At2g45180"
/note="F4L23.31"
4320..4724
/gene="At2g45180"
/codon_start=1
/product="putative proline-rich protein"
/protein_id="AAB82643.1"
/db_xref="GI:2583134"
/translation="MASKALAVTALLITLNLFFFTVSTKCPPTTKPKPKPKPKPKK
APAVKPTCDTTLKGLGVCADLLGLVNVVVGSPKPTCCITLLOGLANLEAAVCLTALK
ANVIGLNLNVPIDITLLNLYCGKKVPHGQCS"
5379..5403
/rpt_family="POLY A"
/complement(join(<6393..6470,6586..6663,7082..7157,
7248..7296,7872..8004,8099..8260,8382..8495))
/gene="At2g45190"
/complement(<6393..>8495)
/gene="At2g45190"
/note="F4L23.30"
/complement(join(6393..6470,6586..6663,7082..7157,
7248..7296,7872..8004,8099..8260,8382..8495))
/gene="At2g45190"
/note="unknown protein"
/codon_start=1
/protein_id="AAB82644.1"
/db_xref="GI:2583135"
/translation="MNSMSSSPSSAVCSVDHSPDHLCYVQCNCQTILAVNVPTT
SLKFTVTRCGCTTLLSNMRSYVLPSANQLQLGPHSYFNPDIIELRDAPSNM
NMNMNGHTMNDIPSPMDLHQOHEIPKAPVNRPEKQRPVSAYNRFKEIQRIK
AGNPDIHSREAFSAAKANWAHFIHFGLVPDNPQVKKTNMPOEGEDNMVKEGFA
PAAANVGVTPT"
10147..10201
/rpt_family="(GA)n"
12613..12646
/rpt_family="(CATA)n"
/complement(join(15613..15729,15821..15950,16066..16137,
16252..16341,16810..17006,17451..17647))
/gene="At2g45200"
/complement(15613..17647)
/gene="At2g45200"
/note="F4L23.29"
/complement(join(15632..15729,15821..15950,16066..16137,
16252..16341,16810..17006,17451..17583))
/gene="At2g45200"
/codon_start=1
/product="putative cis-Golgi SNARE protein"
/protein_id="AAB82642.1"

```

FEATURES
source

/cultivar="Columbia"
/db_xref="taxon:3702"
/clone="F6N15"
/chromosome="IV"
/map="unknown"
/complement(829..4226)
/gene="F6N15.14"
/complement(join(829..942,1006..1086,1156..1237,1315..1500,1574..1675,1776..1809,1863..2001,2106..2230,2325..3299,3548..3709,3844..3987,4073..4226))
/gene="F6N15.14"
/note="contains similarity to breast cancer susceptibility (Brca2)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19315.1"
/db_xref="GI:3193333"
/translation="MSTWLPDPSGDFRWEVAGRIQLQSVSDPTPTKALESTAPLPS
MDLLIQGSKLIAEAMPGEIPMFRTGLGKSVYLKESIAKASILAIEKVITYSDLR
NATCSIQMROVDTAETLPMFTASGKSVPLKESIAKASILGSDKLIIDSNVLPRE
SGGVNSLFTVASKNKVNVSSAGLAKAKALLGLEDDLNGFNHYQNSSSSQOHWGS
GLKTHEEDATVYKHSCTPGQYEDYVSKRSEVLNPSLKVPPTKEQTAGKSKLSVSA
EALKARNLGPDELGPEDDVGAGQFTPEKDERLSIAINNGSANGRYIAHEEKT
SNKHTPSVFLWSSKQFSSVNLNENLASGNLKKKFDAAVDETDALNATHGLSN
RLASDMYNNKSNVGFIPRGQGRPADQPLVDITNRRDTAYANKQDSQKKRLGK
TVSVSPKRPRTISSEKTPSKHALQASGLSVVSCDILT SKVLSTRYPEKSPRYIK
DFGHPKATRTTMEESNOVMQINNYVMSLQTRLEKLFKQWL SILVLPYNNHPE
KVCORSFEACWLYWKLACYDIYPAKCGNFELITNVEELKYRYEREVNHGCSA
IKRILSGPASPMMVLCISAINPKTINDSQAHSGLSCSNVVKVLTQVSMNHALD
VITGLNAGKLVQGLKHAYVLISYIKLIGLAGSGWATPTSPLESFIQPNFALVS
STICLLNTINGYRAHWADRLGCK"
7427..8228
/gene="F6N15.13"
/join(7427..7890,7976..8041,8120..8228)
/gene="F6N15.13"
/note="coded for by A. thaliana cDNA W43412"
/codon_start=1
/protein_id="AAC19310.1"
/db_xref="GI:3193328"
/translation="MALALSACSPPLRTTRAGFRSCSIFANPAORAKRKLLELI
SEDRGLRTQKPKRDEVLNAIESMTVIGRSITTDLSLSTWRLTWTEKEQLEFI
EKAGLFTAGDVLQVDVNNKRLNNVITFPDGPVGFVSDIDIASPQRVNFNSAV
LRGNWELPLPPFGKGFENYMDGEIRVAKDIRGDIYLVDRAPYNWTESF"
8515..9759
/gene="F6N15.12"
/join(8515..8695,8783..9759)
/gene="F6N15.12"
/note="similar to plant chalcone and stilbene synthases"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19299.1"
/db_xref="GI:3193317"
/translation="MLVSRVEKQKRVAYQKATVLALGKALPSNVVQENLVEYLR
EIKDNLSIKDLQHKSTTKYTRYVMSRETLHKYPELATGESPTIKRLEIANDA
VQVMEAYSLVLIKEMGRAVEDITHLVYSSSEFRLPGDLYLSAQLGSLNEVQVRL
YFLGCGSLGRVAKDIAENNPGRSVLLTSETTVLGRPPNKARPNLYGAALFQV
GAAALIGADPTSESPFELHCAVQPLQFQVIGDRLSEEGITFKLGRDLPOKIE
DNVEFCVKLVAKAGSAGLELNDLFWAHPGPAILSGLETKLKLPKLEKRRALM
DYGNVSSNTIFYIMDKVRDELEKKGTEGEWGLGIAFGDITFEGLMRLN"
11756..11756
/gene="F6N15.11"
/join(11756..11816,11897..12294,12667..12732,13158..13535,13624..13710)
/gene="F6N15.11"
/note="contains similarity to transcriptional activators such as Ra-like and myc-like regulatory R proteins"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19308.1"
/db_xref="GI:3193326"
/translation="MLRQPPPPPPAPPQPTSPYVEALTWENGLGLHGLGPPRV
TASSTKYTGAGTLESIVDQATRLPNPKRDELVPFWHRRSSRAMDALVPCSNL
VHEQSKPGVGVSTRVSGSDGRTMGGRKARVAPWEGSGGSQLTMDTYDKRRDKIN

gene
CDS
/gene="F6N15.10"
/join(15299..15951,16806..17061,17136..17246,17320..19104,19156..19234,19332..19628,19799..20001,20363..20614,20746..20892)
/gene="F6N15.10"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19309.1"
/db_xref="GI:3193327"
/translation="MENRHLPSKFFSSASSMAQNQLDLSLTSHISLYHSHSSSMAN
TIPNPSAIRWFSSLSHORLSHLTVDPKFVQILLOMLGYIRTKGPCSFILPDLPL
SSDPLSLCKKSRGLISRVSESESERFVDFSTRFLGSGEGRAODGCSVNSLDSV
VMAEFITTVDRFVETMDVLDGAFNGESDLGSNNWELEKRAKGYISMEAFVANS
SMGTITTVLCILKLRGLFVMSIDVSSELDLNDTHKCSSSSNQKLGSTNRKQK
KTRNMKPTPEAKSDKNVNLKNGKDKQAKLEFNKREAIIECKVPTASTMINDEP
ASATMBVPLGVARKRTKKRKNKSKCTSLNNGEVNKSVMYSAIVKASKCD
SSCTSAHQPOEYINAOIIEHGSFCERNRSGTCASVNGAANCEYSGEESHKAEET
SVTSSDLSVDYPAGPSCENVNPKSCCRDRKEKLTMPNERSRTLDGESHRIHQH
REAGYFASSSSEFVSTEWPAVAPMTFSHVSHLPTATDRHLDDVGHNLHPYVRQFV
STVQHARNPIEGSHQVLSMPMSLDMPVHNSGGLTAFATCNDYDLSILDIPEQ
KNKHELNECENNMLEDEFEVHTVSGVDYNOYFGGVMYTAFNSDLGTGFSRPPLS
SDSSSWAHAEAMKRSVDDMFVSSSYANGLDSPTAASFCSPHLPNQLGYV
PGNEISPKILQAPTTIEGAGEEVSSTLASLSDVEGNSDLSLPILRPIIIPNMS
KSEYKSDYDKSPNPTREHPRIKPPSPVLCVPRAPRPPSPVPSNRRAGFP
TVRSQSSPRHGRHGFHGVWEPRGAEIVLPRNKS LAYRPIQLPGLALLQDH
LIAMSGSRQHEHVSNNVDIYDVAFPLPPELLNCPMGESLHENTMSEIFECI
YFSLWDGVSSPLNAYILLAKVAENMARKPIINWIKRVTSLQVLRVSTNIFGSSA
TGLSLPSSDVLVCLPPVRNLVSKPKIIPHLVKIYISLLKVVVYVNDOSVLFKYI
FYLLMNTCLYIFVLGASLIIPYWLFFLCGCGSHVYIEPIIMLVVPEVPCDLCGI
QSPKDPDCTVDQDSNGTPEMGEFSAANSIPTNTGNLAIAKAVSLDIDSPKTPSH
TGLQTTOLVLDLQTEQFPAAPLAIVLKQFLADRLDQSYSGGLSSCLVRCILLLK"
complement(23534..25288)
/gene="F6N15.15"
complement(join(23534..23629,23711..23845,23934..23999,24049..24092,24457..24490,24753..24823,25132..25288))
/gene="F6N15.15"
/note="contains similarity to Lotus japonicus RING-finger protein (DB:236750)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19316.1"
/db_xref="GI:3193334"
/translation="MSLLSENEFELDLNEIPEKTVYSDGDEFLDNKIPREETV
NNSDEDFADHLHRRNQATKRRKKLRKPRVFRHIEKVISFSLVLRASSFLDLNLW
ADKTDTPYQELRMDTDHMTYEQLLQLCNNGYENSGVKASINDRCLRNTKPSFOSL
ADKTCICQDGFQKRGVGLNGCHNFHNCVKRPWILTCK"
26808..27437
/gene="F6N15.9"
26808..27437
/gene="F6N15.9"
/note="contains similarity to pectinesterases"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC19307.1"
/db_xref="GI:3193325"
/translation="MAANKNLFVLLSLFLIIFSATATSKDYDTKAYVHSCRTTL
YPKLCVSRMSRYVSRVQNPDLARFALKASYRAKYTKAFLLKEVKNLETLRPOY
TASVLDQDITRDSVNLQSLAIELDRVSRQKSGDLHWHNNLTQWTSTALTDAE
YCVSQFGRMSKLKATIKGVKNVEETSNALAFIEHYAARARRP"
27662..27734
/note="codon recognized: AAG; Lys"
/product="tRNA-OTHER"
28096..30456

Query Match 35.2%; Score 135.6; DB 8; Length 95643;
Best Local Similarity 63.9%; Pred. No. 1.5e-16;
Matches 221; Conservative 0; Mismatches 124; Indels 1; Gaps 1;
QY 9 accaagcctgacatcttggttgccaccgctccgagctggagatcgcgggcgcaatgcyg 68

intron STICLLINNGTYRAHWADRLGFCK"
complement(6081..6143)
/number=1
exon complement(6144..6224)
/gene="AT4g00020"
/number=2
intron complement(6225..6293)
/number=2
exon complement(6294..6375)
/gene="AT4g00020"
/number=3
intron complement(6376..6452)
/number=3
exon complement(6453..6638)
/gene="AT4g00020"
/number=4
intron complement(6639..6711)
/number=4
exon complement(6712..6813)
/gene="AT4g00020"
/number=5
intron complement(6814..6913)
/number=5
exon complement(6914..6947)
/gene="AT4g00020"
/number=6
intron complement(6948..7000)
/number=6
exon complement(7001..7139)
/gene="AT4g00020"
/number=7
intron complement(7140..7243)
/number=7
exon complement(7244..7368)
/gene="AT4g00020"
/number=8
intron complement(7369..7462)
/number=8
exon complement(7463..8437)
/gene="AT4g00020"
/number=9
intron complement(8438..8685)
/number=9
exon complement(8686..8847)
/gene="AT4g00020"
/number=10
intron complement(8848..8981)
/number=10
exon complement(8982..9125)
/gene="AT4g00020"
/number=11
intron complement(9126..9210)
/number=11
exon complement(9211..9364)
/gene="AT4g00020"
/number=12
gene 12565..13366
exon 12565..13028
/gene="AT4g00030"
/number=1
CDS join(12565..13028,13114..13179,13258..13366)
/gene="AT4g00030"
notes="coded for by A. thaliana cdna W43412
contains EST gb:W43412"
/codon_start=1
/product="predicted protein of unknown function"
/protein_id="CAB80761.1"
/db_xref="GI:7267090"

/translation="MALALSLSACSPPLRRTRAGPRTSCSIFANPAORAKRKLLELI
SEDRGLRTQDKPKRDEIVNAESMTVIGRSSITDDSDLSATWRLTWTEKEOLFII
EKAGLFGTTAGVDVQVYDYNKRILNNVITFPDGVFVFRSDIDIASPQVNFNSAV
LRGNWELPLPPFGKGVFENVYMDGEIRVAKDIRGDIYLIVDRAPYNWTFVF"

intron 13029..13113
/gene="AT4g00030"
/number=1
exon 13114..13179
/gene="AT4g00030"
/number=2
intron 13180..13257
Query Match 35.2%; Score 135.6; DB 8; Length 197119;
Best Local Similarity 63.9%; Pred. No. 1.3e-16;
Matches 221; Conservative 0; Mismatches 124; Indels 1; Gaps 1;
QY 9 accaagcctgacatcattggttggccaccggctcccgagctggagatcgcgggcaatgcg 68
Db 40211 ACAAAACCTAGACGTAATCCTAATGGGGACAGGTCGGAGCTAGAGATAGCTGCAAGGCT 40152
QY 69 gccacgagctgaggaaggggaagacggtccgcgctcgtctcgttcgtctcctctgggaa 128
Db 40151 GGAGAGAAAGCTAAGAGAGAAAGAAAGACAGTGAGAGTAGTGTCTATTGGTGAGTTGGGAG 40092
QY 129 ctctttgatgagcagtcggtatgatacaagagagagcgtccctccctgccgacgtcacagcg 188
Db 40091 CTGTTTGATGACAAAGAGAGAGAAATACAAAGAAAGTGTTTACCTTGAGAGGTTTCAGCA 40032
QY 189 aggatcagcatcgagggccgggtccactctcgtcgtgaggaagtagctcgaggagcccaagggc 248
Db 40031 AGACTGAGCATTTGAAGCAGGTCGACCTTTGGTGGGAGAGATGGTTGGCTCCCAAGA 39972
QY 249 aaggccattggcatcgacaaagttcggcgagtgctcctcgcgggagacgtatcacaaggag 308
Db 39971 AAAGCCATCGGTGTAGAC-AGTTTGGAGCAAGTGCCAGCAATGTGCTCTACAGGAAA 39913
QY 309 tacggcatcacccgtggagagcatcatctgaactgtccaagagagctttt 354
Db 39912 TTGGACTCACCGTGGATACGGTTATTGACGCCGCTAAGACTTTT 39867

Search completed: November 4, 2000, 13:39:32
Job time: 17421 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:31 ; Search time 320.8 Seconds
(without alignments)
450.842 Million cell updates/sec

Title: US-09-300-482-356

Perfect score: 385

Sequence: 1 caacggcaccacgactgac.....acggtctggagttttttt 385

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36.*

- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	170.4	44.3	2629	17	T35903	Sequence encoding
2	80.6	20.9	11443	19	V52182	Streptococcus pneu
3	80	20.8	1953	18	V74685	Staphylococcus aur
4	63	16.4	1738	19	T98569	DNA encoding a S.
5	55.8	14.5	3225	20	X13233	Enterococcus faeca
6	53.8	14.0	946	19	V52372	Streptococcus pneu
7	51.6	13.4	1872	17	T28491	P. aeruginosa dete
8	44	11.4	1284	21	Z46836	Heat-resistant iso
9	44	11.4	1284	21	Z57439	Heat-resistant iso
10	43	11.2	1281	21	Z46835	Heat-resistant iso
11	43	11.2	1281	21	Z57438	Heat-resistant iso
12	43	11.2	1480	19	V35659	Heat-resistant iso

13	43	11.2	2872	21	Z46834	Heat-resistant iso
14	43	11.2	2872	21	Z57437	Heat-resistant iso
15	41.2	10.7	4195	17	T12068	FLT4 receptor tyro
16	41.2	10.7	4425	16	T03090	Protein tyrosine-k
17	41.2	10.7	9108	16	T03104	Plasmid prk5.tki-1
c 18	41	10.6	1693	21	Z56005	Contig 004 from co
19	39	10.1	1380	20	X22117	Human secreted pro
20	37.2	9.7	723	15	Q73597	Dactylis glomerata
21	36.8	9.6	1008	21	A27338	Rice RA8 anther-sp
22	36.8	9.6	1910	19	V48148	Nicotianamine amin
23	36.8	9.5	3003	21	A27333	Rice RA8 anther-sp
c 24	36.6	9.4	5821	18	T58686	Human lung tumour
25	35.6	9.2	1136	16	O89541	DNA encoding S. fr
26	35.6	9.2	1431	16	O85641	Wheat cold toleran
27	35.6	9.2	1632	14	Q42976	Acetyl-CoA-acyltra
28	35.6	9.2	4980	14	Q36660	Carinomycin 4-O-m
c 29	35.4	9.2	68750	21	Z55887	Sequence of the po
30	35.4	9.1	77536	21	A14651	Sorangium cellulos
31	35	9.0	1026	20	Z28362	Nucleotide sequenc
32	34.8	9.0	1918	19	V68515	Platelet factor-4
33	34.8	9.0	2238	19	V24017	STRL33 coding sequ
34	34.6	9.0	1157	12	Q13716	Human HMBU14 codi
35	34.6	9.0	1157	17	T40789	Geranylgeranyl pyr
36	34.6	9.0	1157	17	T41741	Geranylgeranyl pyr
37	34.6	9.0	1157	17	T41742	Recombinant gerany
38	34.6	9.0	1157	17	T37091	Wild type geranyig
39	34.6	9.0	1157	17	T37092	N-terminally trunc
40	34.6	9.0	1157	17	T37092	Geranylgeranyl pyr
41	34.6	9.0	1157	17	T37092	E. herbicola heran
42	34.6	9.0	1157	18	T91541	Geranylgeranyl pyr
43	34.6	9.0	1157	18	T91542	Erwinia herbicola
44	34.6	9.0	12905	21	Z99101	S. fradiae tylosin

ALIGNMENTS

RESULT 1	
T35903	T35903 standard; cDNA; 2629 BP.
ID	T35903 standard; cDNA; 2629 BP.
XX	
AC	T35903;
XX	
DT	29-JAN-1997 (first entry)
XX	
XX	Sequence encoding tobacco transketolase.
DE	
XX	Transketolase; herbicide; Nicotiana tabacum; Identification;
KW	screening; SS.
KW	
XX	Nicotiana tabacum.
OS	
XX	
FH	Key Location/Qualifiers
FT	60..2291
FT	/tag= a
FT	/product= Transketolase.
XX	
PN	EP723017-A2.
XX	
PD	24-JUL-1996.
XX	
PF	13-JAN-1996; 96EP-0100458.
XX	
PR	23-JAN-1995; 95DE-4001906.
XX	
PA	(SCHM/) SCHMIDT R.
PA	(BADI) BASF AG.
XX	
PI	Schmidt R, Sonnwald U, Stitt M;
XX	
XX	WPI; 1996-335478/34.
DR	P-PSDB; W03319.

[illegible]

progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or other surgical techniques.

XX
sequence 1738 BP: 493 A: 410 C: 365 G: 469 T: 1 other;

Query Match 16.4%; Score 63; DB 19; Length 1738;
Best Local Similarity 51.8%;
Pred. No. 3.4e-07;
Matches 141; Conservative 0; Mismatches 131; Indels

17 tgacatcattgggttgggcaccgctccgagctggagatcggggcaatcggccgacga 76
||| ||| | | | | | | | | | | | |
222 TGATACTATTATCATTTGCTACAGGACTGAGGTCAATCTAGCTATCAAAGTGGTAAGGA 213

Qy 77 gctgaggaaggaggaaacgcgtccgcgtcgtctcgttcctcctgggaactctttga 136
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
d4 212 ammcmmmmmtacacacctgcctaagaatcacctcgctattatgccctcaacgcgaacttttga 153

QY 137 t g a g c a g t c g g a t g a g t a c a a g g a g a g c g t c t c c t g c c g a c g t c a c a g c g a g g a t c a g 196

Qy 197 catcagggccgggtccactctcgtggtggcagaagtaogtcggagcccaaggcgaagcccat 256

Qy 257 tggcatgcacaagtctggcgagtgctctg 288

RESULT 5

AI3233
ID X13233 standard; DNA; 3225 BP.
XX
XX v13233.

DT 19-MAR-1999 (first entry)

XX Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.
XX
PN W09850555-A2.

PD	12-NOV-1998.
XX	
PF	04-MAY-1998; 98WO-US08985.

AA		
PR	14-NOV-1997;	97US-0066009.
PR	06-MAY-1997;	97US-0044031.
PP	16-MAY-1997;	97US-0046655.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XXXXXX

XX
DR
XX
WPI; 1999-045171/04.

PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.

ps Claim 1; Page 1347-1348; 2084pp; English.
xx
cc A computer readable medium has been developed which has recorded on it

CC X12938 to X13919 represent these nucleotide sequences which are primarily
CC X12938 to X13919 represent these nucleotide sequences which are primarily
CC nucleotide sequences, also known as contigs. The computer-based system

can identify fragments of the *Enterococcus faecalis* genome with commercial importance. The products can be used to detect the presence of *Enterococcus faecalis* in samples. They can also be used for diagnosing *Enterococcal* infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, *in vivo* or *in vitro*. In particular the polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an *Enterococcal* infection.

Sequence 3225 BP: 1059 A; 613 C; 695 G; 856 T; 2 other; XX
S0

Query Match 14.5%; Score 55.8; DB 20; Length 3225;
Best Local Similarity 51.0%; Pred. No. 2.5e-05;
Matches 132; Conservative 0; Mismatches 127; Indels 0;

	Qy	24	attgggttgggcacgagctcgaagtggagatcgccggccaatcgccgacgagctgagg	83
	b6	1555	+++++ttccccacttggttcraaatcaatctaccggtgaagcacaaagcgaaatttagcg	1624

[illegible]

Qy 144 tcgatatgatacaaggagagcgtccctcccgacgtcacagcgaggatcagaatcgag 203

Qy 204 g c g g g g t c c a c t c t c g g c t g g c a g a a g t a c g t c g g a g c c c a a g g c c a a g g c c a t t g g c a t c 263

QY 264 gacaagttcggcgcgagtg 282
||| | ||||| |||

DD 1803 gaccacccggcgccccg 1022

RESULT 6

V52372
ID V52372 standard; DNA; 946 BP.

AC V52372;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:239.

... Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
KW computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

AA
PN
WO9818931-A2.

07-MAY-1998.

30-OCT-1997; 97WO-US19588.

PR 31-OCT-1996; 96US-0029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;

WPI: 1998-272225/24.

Computer-readable medium with recorded Streptococcus pneumoniae polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae

XX PS Claim 1; Page 1259-1260; 1409pp; English.

CC The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded

CC on it, or a representative fragment or a sequence at least 95% identical

CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1

CC to 391 (V52134 to V52524) are genomic fragments from *Streptococcus*

CC pneumoniae. The present invention also describes an isolated nucleic acid

CC molecule encoding a homologue of any of the fragments of the *S. pneumoniae*

CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced

CC by a process comprising: (a) screening a genomic DNA library using as a

CC probe a target sequence defined by any of the sequences in SEQ ID NO:1

CC to 391, identifying members of the library which contain sequences

CC that hybridise to the target sequence and isolating the nucleic acid

CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced

CC from an organism, amplifying nucleic acid molecules whose nucleotide

CC sequence is homologous to amplification primers derived from the

CC fragment of the *S. pneumoniae* genome to prime the amplification and

CC isolating the amplified sequences. The computer readable medium can be

CC used in a computer-based system for identifying fragments of the

CC *S. pneumoniae* genome of commercial importance, or expression modulating

CC fragments of the *S. pneumoniae* genome. Products from the present

CC invention can be used in diagnosis kits and assays, and pharmaceutical

CC compositions and vaccines for *S. pneumoniae*.

XX SQ Sequence 946 BP; 272 A; 189 C; 218 G; 267 T; 0 other;

Query Match 14.08; Score 53.8; DB 19; Length 946;

Best Local Similarity 49.08; Pred. No. 6.1e-05;

Matches 171; Conservative 0; Mismatches 177; Indels 1; Gaps 1;

Qy 10 ccaagcctgacatcatgtgggtggccaccggtccgagctggagatcgcgaggcaatgcgg 69

Db 491 ccgacttggatccatcttggatcgacaggttcagaggttaattctgtgtctcagctg 550

Qy 70 ccgacgagctgaggaaggaggagcgtccgctgctgtctgtctgtctcctgggaac 129

Db 551 ctgaagaattgctagtcagagcgcaaaaatccgctagtcagcatcctacagatg 610

Qy 130 tctttgatgacgctgagtagtacaaagagagcgtctcctcctgcgcagctcacagcga 189

Db 611 tctttgatacaagatgcagcttacaaagagaattcttccaaatgcagtcgcgcgc 670

Qy 190 ggatcagcatcagcgcggtccactcgtcgtggcagagagcgtcctcctgcgcagcgaaggca 249

Db 671 gtgtgagtcgaatgggtgcaagtcacaaactgggtacaaatgtgtgtctcgtatgtg 730

Qy 250 aggcattggcagacagaagtgcgcgagtgctcctcgtcgcggagcatctacaaaggagt 309

Db 731 ccgtctaggtattga-tacttcggagcctctgcccagcaaccaaaagtattggcagaat 789

Qy 310 acggcatcaccgtggagagcatcatcgaactgccaaagagcttttaaga 358

Db 790 atggcttactgtagaaatctgtgtaaaagtgtgtcgaactggaata 838

RESULT 7

T28491

ID T28491 standard; DNA; 1872 BP.

XX AC T28491;

XX DT 01-APR-1997 (first entry)

XX DE

XX P. aeruginosa detection probe #2.

XX Detection; probe; amplification primer; bacterial pathogen; pneumonia;

XX *Escherichia coli*; *Klebsiella pneumoniae*; *Pseudomonas aeruginosa*;

XX *Proteus mirabilis*; *Streptococcus pneumoniae*; *Staphylococcus aureus*;

XX *Staphylococcus epidermidis*; *Enterococcus faecalis*; respiratory tract;

XX *Staphylococcus saprophyticus*; *Streptococcus pyogenes*; urinary tract;

KW Haemophilus influenzae; Moraxella catarrhalis; septicaemia; meningitis;

KW infection; intra-abdominal infection; skin infection;

KW bacterial resistance; beta-lactam antibiotic; ds.

OS Synthetic.

XX WO9608582-A2.

XX 21-MAR-1996.

XX 12-SEP-1995; 95WO-CA00528.

XX 12-SEP-1994; 94US-0304732.

XX (BERG/) BERGERON M G.

XX (OUEL/) OUELLETTE M.

XX (ROYP/) ROY P H.

XX Bergeron MG, Ouellette M, Roy PH;

DR WPI; 1996-179953/18.

Method for the detection of bacterial species using probes and primers - allows detection and quantification of antibiotic resistant bacteria in patients, the environment and food

Claim 34; Page 71-72; 216pp; English.

CC The sequences given in T28490-503 represent probes which were used in the method of the invention for the detection of *P. aeruginosa* in a sample. The method comprises using probes and/or amplification primers which are specific, ubiquitous and sensitive for determining the presence and/or amount of nucleic acids from selected bacterial species in any sample, where the bacterial nucleic acid comprises a selected target region hybridisable with the probes or primers. The method comprises contacting the sample with the probes or primers and detecting the presence and/or amount of hybridised primers or amplification products as an indication of the presence and/or amount of the bacterial species. This method may be used to detect commonly encountered bacterial pathogens, e.g. *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Proteus mirabilis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Enterococcus faecalis*, *Staphylococcus saprophyticus*, *Streptococcus pyogenes*, *Haemophilus influenzae* and *Moraxella catarrhalis*. These bacterial species are associated with approx. 90% of urinary tract infections and with a high percentage of other severe infections including septicaemia, meningitis, pneumonia, intra-abdominal infections, skin infections and other severe respiratory tract infections. The method may also be used to evaluate a bacterial resistance to beta-lactam antibiotics.

XX SQ Sequence 1872 BP; 402 A; 651 C; 555 G; 264 T; 0 other;

Query Match 13.48; Score 51.6; DB 17; Length 1872;

Best Local Similarity 47.38; Pred. No. 0.00025;

Matches 156; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

Qy 4 ccggcaccagctgacatcatgttggtggcaccgctccgagctcgagatcgcgagca 63

Db 716 ccgttcctcgccgctacctgctgtgtgctgcgcgcgagcgcggtatcgatcgacc 775

Qy 64 atcgccgcagcagctgaggaaggagggaagcgggtccgctgctgtctctctct 123

Db 776 tgtgggccaagagccgaccaggtcgccgagatcaagatcggtcgcgcgcgcgcg 835

Qy 124 gggaaactctttgatgagcagtcgagatcaagagagcgtctcctccagcagca 183

Db 836 tggaaagctcaagttcaagggtctacgagggaccgctacacccatcgccggtctactggc 895

Qy 184 cagcgagatcagcatgagggccgggttcactctcgtcgtgcaggaagtagctggagccc 243

Db 896 cgcgcagttcgatcatgagcggcgagacccttggaacgaagcagatcgtctccacc 955

```

Query Match      11.4%; Score 44; DB 21; Length 1284;
Best Local Similarity 46.4%; Pred.No. 0.019;
Matches 143; Conservative 0; Mismatches 165; Indels 0; Gaps
Qy 3 acggcgaccacaagcgtcatcatcattgggtgggcacggcgtccgagctcgagatcgagatcgggc 62
973 atcgccccggggccaaatcatcaactacaagacgggcgcacgcgctctttgagggccaccac 103

```

QY	3	accggaaccaagcctgacatcatttggttgggcaccggctccagactgagatccgagc	62
Db	973	atgcgcccccgggccaatactactaacagacggccacgcgtctttgggccaaccac	103
QY	63	aatgcggccgacagcttgagggaagggaagacggttcgcggtcgtctcgttctctcc	122
Db	1033	ggacacgcccccaagtacgctggccaggaacaggtgaaaccagacgctcatctctcc	109
QY	123	tgggaactctttgatgacgcagtcgagatgagataaaggagagcgtctcctccgcagatc	182

CC	isolated from Thermus aquaticus.	
XX		
SQ	Sequence 1281 BP; 264 A; 425 C; 406 G; 186 T; 0 other;	
	Query Match 11.2%; Score 43; DB 21; Length 1281;	
	Best Local Similarity 46.3%; Pred. No. 0.033;	
	Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;	
QY	3 accggcaccgaagctgacatcattgggttggtggcaccgggtcccgagctggagatcgcgggc 62	
Db	973 atgcgccccggggccaacatacaactacaagacggcgccacgctcttttgaggccaccac 1032	
QY	63 aatgcggcgacgagctgaggaagggggaagacggtccgctcgtctcgtctcgtctctcc 122	
Db	1033 ggcaccgcccccaagtacgtggcaggacaaagtgaacccagcagcgtcctctctcc 1092	
QY	123 tgggaactctttgatgacagctgagatgagtaacaagagagcgctccctccctgcgcagctc 182	
Db	1093 ggggagatgatctcgtcctacctggctgggaacgagcgcgccgacctcactcactcagggcc 1152	
QY	183 acagcgagatcagcatcgagcgccgggtccactctcgtggtggcagaagtacgtcggagcc 242	
Db	1153 atggagagaccatcagcaaggcctgtgtcactacgacttccacgcctcctggtggcc 1212	
QY	243 caaggcaaggccattggcgcacaaagttcgcgcgagtgctcctcgcgggacgatctac 302	
Db	1213 gagggcaagccgcacgcttcttaagaccagcgagttcggccaggccctgatccagcac 1272	
QY	303 aaggagt 309	
Db	1273 atggact 1279	
RESULT 12		
V35659		
ID	V35659 standard; DNA; 1480 BP.	
XX		
AC	V35659;	
DT	08-SEP-1998 (first entry)	
XX		
DE	Heat-resistant isocitrate dehydrogenase encoding DNA.	
XX		
KW	Isocitrate dehydrogenase; heat-resistant; thermotable; recombinant;	
KW	Thermus aquaticus; ss.	
OS	Thermus aquaticus.	
XX		
FH	Location/Qualifiers	
FT	180..1460	
FT	/*tag= a	
FT	/product= "isocitrate dehydrogenase"	
XX		
PN	JP10165181-A.	
XX		
PD	23-JUN-1998.	
XX		
PF	09-DEC-1996; 96JP-0328400.	
XX		
PR	09-DEC-1996; 96JP-0328400.	
XX		
PA	(AGEN) AGENCY OF IND SCI & TECHNOLOGY.	
XX		
DR	WPI; 1998-406106/35.	
DR	P-PSDB; W60606.	
XX		
PT	Heat-resistant iso:citrate dehydrogenase gene - may be prepared by	
PT	inserting it into expression vector and transforming it into host	
PT	cell	
XX		
PS	Claim 2; Pages 6-8; 9pp; Japanese.	
XX		
CC	This DNA encodes a heat-resistant (thermostable) isocitrate	
CC	dehydrogenase. A host transformed with a recombinant vector comprising	
CC	the thermostable isocitrate dehydrogenase encoding DNA can be used to	
CC	produce the protein recombinantly. The method can prepare the heat-	
CC	resistant isocitrate dehydrogenase efficiently.	
XX		
SQ	Sequence 1480 BP; 298 A; 508 C; 459 G; 215 T; 0 other;	
	Query Match 11.2%; Score 43; DB 19; Length 1480;	
	Best Local Similarity 46.3%; Pred. No. 0.034;	
	Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;	
QY	3 accggcaccgaagctgacatcattgggttggtggcaccgggtcccgagctggagatcgcgggc 62	
Db	1152 atgcgccccggggccaacatacaactacaagacggcgccacgctcttttgaggccaccac 1211	
QY	63 aatgcggcgacgagctgaggaagggggaagacggtccgctcgtctcgtctcgtctctcc 122	
Db	1212 ggcaccgcccccaagtacgctggccaggacaaaggtgaacccagcagcgtcctctctcc 1271	
QY	123 tgggaactctttgatgacagctcgtgagatgagtaacaagagagcgctccctcccgacgctc 182	
Db	1272 ggggagatgatcgtctcgtacctggctgggaacgagggcgccgacctcactcactcagggcc 1331	
QY	183 acagcgagatcagcatcgagcgccgggtccactctcgtggtggcagaagtacgtcggagcc 242	
Db	1332 atggagagaccatcagcaaggcctgtgtcactacgacttccacgcctcctggtggcc 1391	
QY	243 caaggcaaggccattggcgcacaaagttcgcgcgagtgctcctcgcgggacgatctac 302	
Db	1392 gagggcaagccgcacgcttcttaagaccagcgagttcggccaggccctgatccagcac 1451	
QY	303 aaggagt 309	
Db	1452 atggact 1458	
RESULT 13		
V46834		
ID	V46834 standard; DNA; 2872 BP.	
XX		
AC	V46834;	
DT	10-APR-2000 (first entry)	
XX		
DE	Heat-resistant isocitrate dehydrogenase encoding genomic DNA.	
XX		
KW	Heat-resistant; isocitrate dehydrogenase; ICD; thermostable;	
KW	medical chemistry; agricultural; pharmacology; ds.	
OS	Thermus aquaticus.	
XX		
PN	JP11346780-A.	
XX		
PD	21-DEC-1999.	
XX		
PF	09-JUN-1998; 98JP-0176629.	
XX		
PR	09-JUN-1998; 98JP-0176629.	
XX		
PA	(AGEN) AGENCY OF IND SCI & TECHNOLOGY.	
XX		
DR	(MIYA/) MIYAZAKI K.	
DR	(ORIY) ORIENTAL YEAST CO LTD.	
XX		
DR	WPI; 2000-109695/10.	
DR	P-PSDB; Y56857.	
XX		
PT	DNA of a structural gene coding a heat-resistant isocitrate	
PT	dehydrogenase - useful as a reagent for research in medical chemistry,	
PT	biochemistry, agricultural chemistry and pharmacology	
XX		
PS	Claim 1; Page 9-12; 22pp; Japanese.	

XX The invention provides a DNA of a structural gene coding a heat-resistant
CC isocitrate dehydrogenase (ICD) starting from initiation codon and ending
CC at terminating codon of the 2872 base sequence (Z46834), where the second
CC codon GCC from the initiation codon ATG is replaced by GCT, GCA or GCG,
CC and/or the ninth codon CCC is replaced by CCT, CCA or CCG. The ICD is
CC useful as a reagent for researches in medical chemistry, biochemistry,
CC agricultural chemistry and pharmacology. The method can prepare a heat-
CC resistant ICD in large amount at low cost. The present sequence
CC represents the genomic DNA encoding ICD.
XX
SQ Sequence 2872 BP; 519 A; 988 C; 966 G; 399 T; 0 other;

Query Match 11.2%; Score 43; DB 21; Length 2872;
Best Local Similarity 46.3%; Pred. No. 0.039;
Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
QY 3 accggcaccagcctgacatcattgggtgggacccgctccgagctggagatcgcgggc 62
DB 1152 atcgccccggggcccaactcaactacaagacggcccgctgttgaggccaccac 1211
QY 63 aatgcgccgacgactgaggagagggggaagacggtcccgctgtctgtctgtctcc 122
DB 1212 ggcacgcgcccaagtacgctggccaggacaaggtgaaccccgagcgctcatctccc 1271
QY 123 tgggaactcttgatgagcagctgagtgacagagagagagcgctccctcgccagcgtc 182
DB 1272 ggggagatgatctctgctacctggctggcagagggcgccgactcatcagggcc 1331
QY 183 acagcgagatcagcatcgagccgggtccactctcgctggcagagtagctcgagacc 242
DB 1332 atggagagacatcagcaaggcgtgttcactacgactccacgcctcctgtgtggcc 1391
QY 243 caaggcaaggccattggtcagtcacagatcgcgagtgctcctgcgggacgatctac 302
DB 1392 gaggcgaccccgccacgcttcttaagaccagcgagttcgccaggccctgatccagcac 1451
QY 303 aaggagt 309
DB 1452 atggact 1458

RESULT 14 ...
ID Z57437 standard; DNA; 2872 BP.
AC Z57437;
XX
XX 10-APR-2000 (first entry)
XX Heat resistant isocitrate dehydrogenase encoding DNA SEQ ID NO:2.
XX Heat resistant; isocitrate dehydrogenase; ICD; ammonia; determination;
XX ds.
XX Thermus aquaticus.
XX
XX Key Location/Qualifiers
XX CDS 180..1460
XX /*tag= a
XX /product= "heat resistant isocitrate dehydrogenase"
XX
XX JP11346781-A.
XX
XX 21-DEC-1999.
XX
XX 09-JUN-1998; 98JP-0176643.
XX
XX 09-JUN-1998; 98JP-0176643.
XX
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
PA (MIYA/) MIYAZAKI K.

PA (ORIY) ORIENTAL YEAST CO LTD.
XX WPI; 2000-109696/10.
DR P-PSDB; Y67969.
XX
XX A liquid reagent for the elimination of ammonia - useful for the
XX determination of biosubstances in a sample
XX
XX Claim 5; Page 15-18; 28pp; Japanese.
XX
XX The present invention describes a solution type reagent for the
XX elimination of ammonia containing a heat-resistant isocitrate
XX dehydrogenase (ICD) showing a high stability at an alkaline pH and
XX responding to both NADP+ and NAD-. The reagent is used for the
XX determination of biosubstances in a sample. The reagent can eliminate
XX ammonia in a very short period. The present sequence encodes ICD
XX isolated from Thermus aquaticus.
XX
SQ Sequence 2872 BP; 519 A; 988 C; 966 G; 399 T; 0 other;

Query Match 11.2%; Score 43; DB 21; Length 2872;
Best Local Similarity 46.3%; Pred. No. 0.039;
Matches 142; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
QY 3 accggcaccagcctgacatcattgggtgggacccgctccgagctggagatcgcgggc 62
DB 1152 atcgccccggggcccaactcaactacaagacggcccgctgttgaggccaccac 1211
QY 63 aatgcgccgacgactgaggagagggggaagacggtcccgctgtctgtctgtctcc 122
DB 1212 ggcacgcgcccaagtacgctggccaggacaaggtgaaccccgagcgctcatctccc 1271
QY 123 tgggaactcttgatgagcagctgagtgacagagagagagcgctccctgcgcagctc 182
DB 1272 ggggagatgatctctgctacctggctggcagagggcgccgactcatcagggcc 1331
QY 183 acagcgagatcagcatcgagccgggtccactctcgctggcagagtagctcgagacc 242
DB 1332 atggagagacatcagcaaggcgtgttcactacgactccacgcctcctgtgtggcc 1391
QY 243 caaggcaaggccattggtcagtcacagatcgcgagtgctcctgcgggacgatctac 302
DB 1392 gaggcgaccccgccacgcttcttaagaccagcgagttcgccaggccctgatccagcac 1451
QY 303 aaggagt 309
DB 1452 atggact 1458

RESULT 15
ID T12068 standard; DNA; 4195 BP.
XX
XX AC T12068;
XX
XX DT 08-APR-1996 (first entry)
XX
XX FLTA receptor tyrosine kinase DNA.
XX
XX DE FLTA4; receptor tyrosine kinase; probe; metastasis; lymphoma;
XX lymphangioma; immunological disease; cancer; diagnosis; therapy; ss.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX CDS 20..3976
XX /*tag= a
XX
XX WO9533772-A1.
XX
XX PN 14-DEC-1995.
XX
XX PD
XX

```
PF 09-JUN-1995; 95WO-FI00337.
XX
PR 09-JUN-1994; 94US-0257754.
XX
XX (ALIT//) ALITALO K.
PA (KAIP//) KAIPAINEN A.
PA (KARN//) KARNANI P.
PA (KORH//) KORHONEN J.
PA (MATI//) MATIKAINEN M.
PA (MUST//) MUSTONEN T.
PA (PAJU//) PAJUSOLA K.
XX
XX Alitalo K, Kaipainen A, Karnani P, Korhonen J, Matikainen M;
PI Mustonen T, Pajusola K;
XX
DR WPI; 1996-040189/04.
DR P-PSDB; R90528.
XX
XX Antibodies against FLT4 receptor tyrosine kinase, hybridomas and
PT nucleic acid probes - used to diagnose and treat e.g. metastatic
PT cancers, involving alterations to lymphatic vessels.
XX
XX Claim 9; Page 27-33; 54pp; English.
XX
XX A DNA sequence (T12068) coding for FLT4 receptor tyrosine kinase
CC (R90528) is used as a probe that specifically binds/hybridizes to
CC DNA encoding human FLT4. Such probes are used to detect FLT4, a
CC novel marker for lymphatic vessels and some high endothelial
CC venules, in biological tissue for use in diagnosis and therapy of
CC e.g. inflammatory, infectious and immunological diseases, metastatic
CC lymph nodes and lymphangiomas.
XX
SQ Sequence 4195 BP; 889 A; 1278 C; 1305 G; 723 T; 0 other;
```

Query Match 10.7%; Score 41.2; DB 17; Length 4195;
Best Local Similarity 46.3%; Pred. NO. 0.12;
Matches 136; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
Qy 51 gagatcgcgggcaatcgccgacgagctgtgaggaaggaggagagagcgctcgcgtc 110
   || || || || || || || || || || || || || || || || || || ||
Db 2075 gacctctgtgacgtgagcgtcgtggtgagatcagtgcttggccggagcgac 2134
   || || || || || || || || || || || || || || || || || || ||
Qy 111 tcgtctctcctcgggaactctttgatgacgagtcgagtgatcacaggagagcgctc 170
   || || || || || || || || || || || || || || || || || || ||
Db 2135 gcgccagcatcgtgtgtacaagacgagagcgtcgtggagaaaagctggagtcgac 2194
   || || || || || || || || || || || || || || || || || || ||
Qy 171 cctgccagctcacagcagagatcagcatcgagggccgggtccactctcggctggcagaag 230
   || || || || || || || || || || || || || || || || || || ||
Db 2195 ttggcgactccaaccagaagctgagcatccagcgcgtgcgagaggatgcgggacgc 2254
   || || || || || || || || || || || || || || || || || || ||
Qy 231 taagtcgagagcccaaggcaagccattggcatcgacaagttcggcgagtgctcctgcc 290
   || || || || || || || || || || || || || || || || || || ||
Db 2255 tatctgtcagcgtgtgtcaacgcaagggtcgtcgtaactctccgccagcggtggccgtg 2314
   || || || || || || || || || || || || || || || || || || ||
Qy 291 gggacgatctacaaggaggtacgcacacgtgtgagagatcatctgcaactgcc 344
   || || || || || || || || || || || || || || || || || || ||
Db 2315 gaaggtccgaggataaggcagcagcagtgagatcgtgatccttgcgtaccggc 2368
   || || || || || || || || || || || || || || || || || || ||
```

Search completed: November 4, 2000, 13:45:37
Job time: 16464 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:15 ; Search time 4352.3 Seconds
(without alignments)
546.925 Million cell updates/sec

Title: US-09-300-482-356
Perfect score: 385
Sequence: 1 caacggcaccagcctgac.....acggtctggaggattttttt 385

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues
Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_est1.*
2:	gb_est2.*
3:	gb_est3.*
4:	gb_est4.*
5:	gb_est5.*
6:	gb_est6.*
7:	gb_est7.*
8:	gb_est8.*
9:	gb_est9.*
10:	gb_est10.*
11:	gb_est11.*
12:	gb_est12.*
13:	gb_est13.*
14:	gb_est14.*
15:	gb_est15.*
16:	gb_est16.*
17:	gb_est17.*
18:	gb_est18.*
19:	gb_est19.*
20:	gb_est20.*
21:	gb_est21.*
22:	gb_est22.*
23:	gb_est23.*
24:	gb_est24.*
25:	gb_est25.*
26:	gb_est26.*
27:	gb_est27.*
28:	gb_est28.*
29:	gb_est29.*
30:	gb_est30.*
31:	gb_est31.*
32:	gb_est32.*
33:	gb_est33.*
34:	gb_est34.*
35:	gb_est35.*
36:	gb_est36.*
37:	gb_est37.*
38:	gb_est38.*
39:	gb_est39.*
40:	gb_est40.*
41:	em_estba.*
42:	em_estfun.*
43:	em_esthum1.*

44:	em_esthum2.*
45:	em_esthum3.*
46:	em_esthum4.*
47:	em_esthum5.*
48:	em_esthum6.*
49:	em_esthum7.*
50:	em_esthum8.*
51:	em_esthum9.*
52:	em_esthum10.*
53:	em_esthum11.*
54:	em_esthum12.*
55:	em_esthum13.*
56:	em_esthum14.*
57:	em_esthum15.*
58:	em_esthum16.*
59:	em_esthum17.*
60:	em_esthum18.*
61:	em_esthum19.*
62:	em_esthum20.*
63:	em_estin1.*
64:	em_estin2.*
65:	em_estin3.*
66:	em_estin4.*
67:	em_estom.*
68:	em_estov1.*
69:	em_estov2.*
70:	em_estpl1.*
71:	em_estpl2.*
72:	em_estpl3.*
73:	em_estpl4.*
74:	em_estpl5.*
75:	em_estro1.*
76:	em_estro2.*
77:	em_estro3.*
78:	em_estro4.*
79:	em_estro5.*
80:	em_estro6.*
81:	em_estro7.*
82:	em_estro8.*
83:	em_estro9.*
84:	em_estro10.*
85:	em_estro11.*
86:	em_estro12.*
87:	em_estro13.*
88:	gb_gss1.*
89:	gb_gss2.*
90:	gb_gss3.*
91:	gb_gss4.*
92:	em_gss1.*
93:	em_gss2.*
94:	em_gss3.*
95:	em_gss4.*
96:	gb_gss5.*
97:	gb_gss6.*
98:	gb_gss7.*
99:	gb_gss8.*
100:	gb_gss9.*
101:	em_gss5.*
102:	em_gss6.*
103:	em_gss7.*
104:	em_gss8.*
105:	em_gss9.*
106:	em_gss10.*
107:	em_gss11.*
108:	gb_gss10.*
109:	gb_gss11.*
110:	em_gss12.*
111:	gb_gss12.*
112:	gb_gss13.*
113:	gb_gss14.*
114:	gb_gss15.*
115:	gb_gss16.*
116:	gb_gss17.*

117: gb_gss18:*
118: gb_gss19:*
119: em_gss13:*
120: gb_gss20:*
121: gb_gss21:*
122: gb_gss22:*
123: gb_gss23:*
124: gb_gss24:*
125: em_gss14:*
126: em_gss15:*
127: em_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	343.6	89.2	541	33	BE129685 945029E06
C 2	331.4	86.1	526	20	AW216325 687049D06
C 3	317	82.3	592	20	AW202478 687037D05
C 4	273.4	71.0	463	11	AI586588 486049B09
C 5	273.4	71.0	562	13	AI855050 606074D04
C 6	273.4	71.0	657	20	AW147017 707009C12
C 7	258.4	67.1	526	24	AW746961 WSL_56_CO
C 8	249.6	64.8	620	35	BE490044 WHE0364_B
C 9	246.4	64.0	581	35	BE364694 P11_14_CO
C 10	246.4	64.0	619	22	AW400161 707052H08
C 11	238.4	61.9	540	24	BE215654 HV_CEB000
C 12	237.2	61.6	705	23	AW565303 LGL_342_G
C 13	235.2	61.1	811	35	BE420889 HWM003_F0
C 14	229.6	59.6	286	35	BE344940 946030B04
C 15	226.6	58.9	747	35	BE418437 SCL033.CO
C 16	224.2	58.2	709	15	AU096596 AU096596
C 17	221	57.4	581	35	BE422449 WHE0055_H
C 18	194	50.4	483	21	AW284966 LGL_296_E
C 19	191.4	49.7	664	12	AI649526 60305G06
C 20	189.4	49.2	476	9	AI163535 A085p38u
C 21	184	47.8	458	24	AW781561 s181905.Y
C 22	183.8	47.7	590	36	C99650 C99650_Rice
C 23	183.4	47.6	646	24	AW775545 EST334610
C 24	183.4	47.6	591	24	AW776336 EST335401
C 25	179.8	46.7	570	13	AI822218 L0_594T3
C 26	178.2	46.3	382	24	AW734032 sk79f08.Y
C 27	175.4	45.6	464	35	BE488527 WHE1058_A
C 28	174.4	45.3	492	18	AV391286 AV391286
C 29	174.4	45.3	531	18	AV388280 AV388280
C 30	171.6	44.6	644	21	AW310464 sf36e03.X
C 31	170.8	44.4	960	33	BE053034 GA_Pa003
C 32	170.6	44.3	573	20	AW203636 sf36e03.Y
C 33	170.2	44.2	518	22	AW508848 s141b02.Y
C 34	170	44.2	484	13	AI900158 sc01f09.Y
C 35	170	44.2	526	20	AW201472 sf03e09.Y
C 36	169.2	43.9	581	14	AI966148 sc34c04.Y
C 37	166.8	43.3	484	24	AW781142 s189f08.Y
C 38	164.4	42.7	551	20	AW156809 se31b02.Y
C 39	163.8	42.5	635	18	AV524053 AV524053
C 40	162.4	42.2	651	12	AI726103 BNLGH1502
C 41	162.2	42.1	563	18	AV523269 AV523269
C 42	162.2	42.1	579	18	AV523749 AV523749
C 43	162.2	42.1	580	18	AV523732 AV523732
C 44	162.2	42.1	583	18	AV523633 AV523633
C 45	162.2	42.1	619	18	AV522729 AV522729

ALIGNMENTS

RESULT 1
BE129685/c

LOCUS	BE129685	541 bp	mrna	EST	21-JUN-2000
DEFINITION	945029E06.X2 945 - Mixed adult tissues from Walbot lab, same as 707 (SK) Zea mays cDNA, mRNA sequence.				
ACCESSION	BE129685				
VERSION	BE129685.1	GI:8577048			
KEYWORDS	EST.				
SOURCE	Zea mays.				
ORGANISM	Zea mays.				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.				
AUTHORS	Walbot.V.				
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 945029 row: E column: 06.				
FEATURES	Location/Qualifiers				
source	1. .541 /organism="Zea mays" /cultivar="W23" /db_xref="taxon:4577" /clone_lib="945" - Mixed adult tissues from Walbot lab, same as 707 (SK) /tissue_type="tassel, kernal, silk, husk, root, leaf" /dev_stage="fully-grown" /lab_host="DH10B" /note="Organ: tassel, kernal, silk, husk, root, leaf; Vector: pGAD10; Site: 1; EcoRI: cDNA library from fully differentiated maize tissues from an active Mutator plant. Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk, root, leaf). Unidirectionally cloned. New library number given to library 707 for additional sequencing."				
BASE COUNT	125 a 169 c 131 g 116 t				
ORIGIN					
Query Match	89.2%	Score 343.6;	DB 33;	Length 541;	
Best Local Similarity	94.9%	Pred. No. le-82;			
Matches	355;	Conservative	0;	Mismatches 19;	Indels 0;
Gaps	0;				
QY	12	aagcctacatcattggttggacaccgctccgagctggagatcgcggaatgcgccc	71		
Db	541	AAGCCTGACCTCATGTGTGATGGGCGCGCTCCGAGTGGAGATCGGGCGGCGCC	482		
QY	72	gacgagctgaggaaggaggagacggttcgctgctctctctctctctctctctctctct	131		
Db	481	GACGAGCTGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	422		
QY	132	tttgatgagcagtcgagatgagatgagatgagatgagatgagatgagatgagatgagatg	191		
Db	421	TTTGATGAGCAGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	362		
QY	192	atcagatcagagc	251		
Db	361	ATCAGCATTTGAGGCTGGGCTCCCTCTTGGCTGGCAGAGTACGTCTGGAGGCCAGGCAAG	302		
QY	252	gccattggatcgacaaagtctcgcgagtgctctctcccgaggacgacatctacaagaatc	311		
Db	301	GCCATTGGATCGACAAAGTTCCGGCGAGTCTCTCTCCGGGAGCATCTCAAGAGGATAC	242		
QY	312	ggcatcaccgtggagagatcatgtgcaatgcccaagagcttttaagagcttaacacggtc	371		
Db	241	GGCATCCCGGGAGGAGCATCATTTGCACCTGCCAAGAGCTTTTAAGAGCTAACACGGTC	182		
QY	372	tgaggagttttttt 385			

Db 322 TGCAGCTGCCAAGAGCTTTTAAGAGCTAAACAGCTCTGAGTTTTTTTCT 272

RESULT 6

AW147017

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .657

/organism="Zea mays"

/cultivar="W23"

/db_xref="taxon:4577"

/clone_lib="707 - Mixed adult tissues from Walbot lab (SK)"

/tissue_type="tassel, kernel, silk, husk, root, leaf"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: tassel, kernel, silk, husk, root, leaf; Vector: pGAD10; Site_1: EcoRI; cDNA library from fully differentiated maize tissues from an active Mutator plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk, husk, root, leaf). Unidirectionally cloned."

BASE COUNT 133 a 203 c 198 g 123 t

ORIGIN

Query Match

Best Local Similarity 71.0%; Score 273.4; DB 20; Length 657;

Matches 280; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 caaccggcacaagctgacatcattgggtggccaccggtccgagctggagatcgcg 60

Db 367 CGACCGCCCAACAGCCTGACCTCATTTGTATGGCCACCGCTCCGAGCTGGAGATCGGG 426

Qy 61 gcaatgcgcccagcagctaggaaaggagggaagcagctccgctcgctctgctgtct 120

Db 427 CCAAGCGCCGCGAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 486

Qy 121 cctgggaactctttgatgagcagctcagatgagtagatgagtagatgagtagatgag 180

Db 487 CCTGGGAACTCTTTGATGACGACCTCGGATGAGTACAGGAGGAGGAGGAGGAG 546

Qy 181 tcacagcagagatcagatcagagcgggtccactctcggtcggtcggtcggtcggtcg 240

Db 547 TCACAGCGAGGATCAGCATCGAGCGCGGTCCACACTCTCGCTGGCAGAGATGACGCGG 606

Qy 241 ccaaggcagagcattgagcagcagcagcagcagcagcagcagcagcagcagcagc 291

Db 607 CCCAGGCGAAGGCGCATGGCATCGACAAGTTCCGGCGGCGAGTGCTCCTCGCG 657

RESULT 7

AW746961

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .526

/organism="Sorghum bicolor"

/db_xref="taxon:4538"

/clone_lib="Water-stressed 1 (WS1)"

/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 134 a 110 c 151 g 131 t

ORIGIN

Query Match

Best Local Similarity 67.1%; Score 258.4; DB 24; Length 526;

Matches 289; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 46 agctggagatcgcgccgaatcgccgagcagcagcagcagcagcagcagcagcagcag 105

Db 1 AACTGGAGATTCGGAAAGGCTGCCGATGAGTTGAGGAAGGAGGAGGAGGAGGAG 60

Qy 106 tcgtctgttcgtct 165

Db 61 TCSTCTCATTTGTTCTTGGAACTTTTCGATCAACAGTCAGATCAATACAGGAGATG 120

Qy 166 tcttccctcgccagctcagcagcagcagcagcagcagcagcagcagcagcagcag 225

Db 121 TCTGCTCTGAGCGCTCAGCGCAAGAAATCAGCATTTGAGCGCGGTCTACTCTCGCTG 180

Qy 226 agaagtcagctcgagcccaagcagcagcagcagcagcagcagcagcagcagcagc 285

Db 181 AGAAGTATGTCGAGAGCCAGGCGAGGCCATTTGGCATCGACAAATTCGGTGGCAGT 240

Qy 286 ctccggagcagatctacaaggagtagcagcagcagcagcagcagcagcagcagcagc 345

Db 241 CTGCCGGAAGGATCTACAGGAGTATGGCATCCTGTGGAGAGCGCTCATTCGGCGGCG 300

Qy 346 agagctttaagagctacaacacggtctggagagtttttttttttttttttttttt 385

Db 301 AGAGCTTTTAAGAGTTAACACAACAGCTCTTGAGTTTTTTT 340

AW746961 526 bp mRNA EST 19-JUL-2000
WSL_56_C06_g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
mRNA sequence.

AW746961
AW746961.1 GI:7660699

EST.
sorghum.

Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.

1 (bases 1 to 526)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
L.H.

An EST database from Sorghum: water-stressed plants
Unpublished (2000)
Contact: Cordonnier-Pratt MM

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860
Fax: 706 542 1805

Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.

Seq primer: T7
High quality sequence start: 73
High quality sequence stop: 526
POLYA=Yes.

Location/Qualifiers
1. .526
/organism="Sorghum bicolor"

/db_xref="taxon:4538"

/clone_lib="Water-stressed 1 (WS1)"

/note="Organ: Mix of 5-week old plants on days 7 & 8 after
water was withheld; Vector: Lambda Zap; Site_1: XhoI;
Site_2: EcoRI; The library was made from poly-A RNA in the
cloning vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."

BASE COUNT 134 a 110 c 151 g 131 t

ORIGIN

Query Match

Best Local Similarity 67.1%; Score 258.4; DB 24; Length 526;

Matches 289; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 46 agctggagatcgcgccgaatcgccgagcagcagcagcagcagcagcagcagcagcag 105

Db 1 AACTGGAGATTCGGAAAGGCTGCCGATGAGTTGAGGAAGGAGGAGGAGGAGGAG 60

Qy 106 tcgtctgttcgtct 165

Db 61 TCSTCTCATTTGTTCTTGGAACTTTTCGATCAACAGTCAGATCAATACAGGAGATG 120

Qy 166 tcttccctcgccagctcagcagcagcagcagcagcagcagcagcagcagcagcag 225

Db 121 TCTGCTCTGAGCGCTCAGCGCAAGAAATCAGCATTTGAGCGCGGTCTACTCTCGCTG 180

Qy 226 agaagtcagctcgagcccaagcagcagcagcagcagcagcagcagcagcagcagc 285

Db 181 AGAAGTATGTCGAGAGCCAGGCGAGGCCATTTGGCATCGACAAATTCGGTGGCAGT 240

Qy 286 ctccggagcagatctacaaggagtagcagcagcagcagcagcagcagcagcagcagc 345

Db 241 CTGCCGGAAGGATCTACAGGAGTATGGCATCCTGTGGAGAGCGCTCATTCGGCGGCG 300

Qy 346 agagctttaagagctacaacacggtctggagagtttttttttttttttttttttt 385

Db 301 AGAGCTTTTAAGAGTTAACACAACAGCTCTTGAGTTTTTTT 340


```

Db 24 CGAGCGAAACAGCCCTGACTACATCAATAGAGCACTGGTTCTGAAGTAGAGATTCGTG 83
QY 61 gcaatgcgcgcagcagctaggaagaggaagcggtcccgctgcgtctctctctctctct 120
Db 84 TGAAGGCTGCGAAGAGTTGACGAGAGGAGGAGGAGCCGTCGCGGTTGTGTCATTGTCT 143
QY 121 cctgggaactctttgatgagcagtcgagtcagtcagtcagtcagtcagtcagtcagtc 180
Db 144 GCTGGGAGCTTTTGGACGAGCAGTCCGATGAATACAGAGAGAGCGTCTCCCTGAGGCTG 203
QY 181 tcacagcagagatcagatcagagcgggtccactctcgcgtcgcagaaagtcagtcgag 240
Db 204 TCACCGCAAGGATCAGCATGAGCTGGCTCTACCCCTCGGATGGCAAAAGTACGTGGGT 263
QY 241 cccaagcagagcagtcgagtcagtcagtcagtcgagtcgagtcgagtcgagtcgagtc 300
Db 264 CCAAGGCGAAGACCATCGCATCGACAAAGTTCGGCGCCAGCGCTCTCTGCCCAAGATCT 323
QY 301 acaaggagtcagcagtcacagtcgagtcagtcagtcagtcagtcagtcagtcagtcag 360
Db 324 ACAAGGAGTACGGATCACTCGCGAGAACGTCATCGCGCCGACGCAAGAGCGCTGAAGAC 383

```

RESULT 12

AW565303 705 bp mRNA EST 19-JUL-2000
 LOCUS LG1_342_G05_g1_A002 Light Grown 1 (LG1) Sorghum bicolor cDNA, mRNA
 DEFINITION

ACCESSION AW565303
 VERSION AW565303.1 GI:7219181
 KEYWORDS EST.
 SOURCE sorghum.

ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Sorghum.
 Magnoliophyta; Liliopsida; Poales; Sorghum.

REFERENCE 1 (bases 1 to 705)

AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.

TITLE An EST database from Sorghum: light-grown seedlings

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: T7
 High quality sequence start: 103
 High quality sequence stop: 705

POLYA=Yes.
 Location/Qualifiers

FEATURES
 source

1..705
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Light Grown 1 (LG1)"

/note="Organ: 10- to 14-day-old light-grown (greenhouse) seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 167 a 139 c 187 g 212 t
 ORIGIN

Query Match 61.6%; Score 237.2; DB 23; Length 705;
 Best Local Similarity 85.9%; Pred. No. 5.4e-54;
 Matches 263; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 80 gaggaaggagggaagcagtcgcgtctctctctctctctctctctctctctctctctctct 139

```

Db 1 GAGGAGAGGGGAGAGACAGTCCGTGCTGCTCATTTTCTTTGGGAACATTTTCGATGA 60
QY 140 gcagtcgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 199
Db 61 ACAGTTCAGATGAATACAAGAGAGAGTGTCTGCTGTCGACCGCTCACCGCAAGATCAGCAT 120
QY 200 cgaagccgggtccactctcgcgtcgcagaaagtcagtcgagtcgagtcgagtcgagtc 259
Db 121 TGAGGCCGGGTCTACTCTCGGCTGGCAGAGTATGTCGGAGCCAGGCGCAGGCGCATGG 180
QY 260 catcgacaagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtc 319
Db 181 CATCGACAAATTCGTCGGAGTGTCTTCCGCGGAACGATCTACAAGAGATATGGCATCAC 240
QY 320 cgtgagagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 379
Db 241 TGTGGAGAGCGTCATTCGGCGGCCCAAGAGCTTTTAAGAGTTAACACACACAGCTTTGAG 300
QY 380 tttttt 385
Db 301 TTTT 306

```

RESULT 13

BE420889 811 bp mRNA EST 24-JUL-2000
 LOCUS HWM003.F02 ITEC HWM Barley Leaf Library Hordeum vulgare cDNA clone
 DEFINITION

ACCESSION HWM003.F02, mRNA sequence.
 VERSION BE420889
 KEYWORDS BE420889.1 GI:9418732
 EST.

SOURCE barley.

ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Poaceae; Hordeum.

Magnoliophyta; Liliopsida; Poales; Hordeum.

REFERENCE 1 (bases 1 to 811)

AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P., Herrmann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y., Pecchioni, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.

TITLE International Triticeae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae

JOURNAL Unpublished (2000)

COMMENT Contact: Herrmann RG
 Botanisches Institut der LMU
 Menzinger Str. 67, D-80638 Munchen GERMANY
 Fax: 49 30 171683
 Email: herrmann@botanik.biologie.uni-muenchen.de
 International Triticeae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.

Location/Qualifiers

FEATURES
 source

1..811
 /organism="Hordeum vulgare"
 /cultivar="Barke"
 /db_xref="taxon:4513"
 /clone_lib="HWM003.F02"
 /clone_lib="ITEC HWM Barley Leaf Library"
 /tissue_type="leaf"
 /dev_stage="14 day old"
 /note="Vector: pBluescriptSK(-); 850 bp average insert size."

BASE COUNT 207 a 199 c 222 g 183 t
 ORIGIN

Query Match 61.1%; Score 235.2; DB 35; Length 811;
 Best Local Similarity 78.3%; Pred. No. 1.9e-53;
 Matches 282; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1 caacggcaccagcctgcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 60

```

Db 165 CGAGTGGAAACAAGCTGACTACATCATATGAGCACGTGTTCTGAAGTAGAGATTGCTG 224
Qy 61 gcaatgcgcgcacagctgaggaaggaggggaaagacgctgcgctctctctcgtct 120
Db 225 TGAAGGCTCCCGAAGAGTTGACGAAGAGGGGGAAGACCGCTCGCGCTTGTCATTGCT 284
Qy 121 cctgggaactcttctgatgagcagctcgatgagatcaagagagagcgtcctccctgcgcag 180
Db 285 GCTGGGAGCTTTTTCACGAGCAGTCCGATGAATACAAGAGAGCGTGTCTCCCTGAGGCTG 344
Qy 181 tcacagcaggatcagcatcgagcggcgggtccactctcgtctggtggcagagtagctcgag 240
Db 345 TCACCGCAAGGATCAGCATTTGAAGCTGCTACCTCGGATGCGCAAGTAGTACGTCGGGT 404
Qy 241 ccaaggaagcgaattggcatcgacagtgctggcgagtgctctcggcgagcatct 300
Db 405 CCAAGGGGAAGACCATCGGCATCGACAAGTTCCGGCGCCAGCGCTCTGCCCCCAAGATCT 464
Qy 301 acaagagtagcggcatcaccgtggagagcatcattgcaactgccaagagcttttaagagc 360
Db 465 ACAAGGAGTAGCGCATCACTGCAGAGAACGTCATCGCGCGCAAGCAAGAGCGCTGAAGACC 524

```

```

RESULT 14
BE344940
LOCUS BE344940 286 bp mRNA EST 17-JUL-2000
DEFINITION 946030B04.y1 946 - tassal primordium prepared by Schmidt lab Zea
ACCESSION BE344940
VERSION BE344940
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.

REFERENCE 1
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946030 row: B column: 04.
Location/Qualifiers
1. .286
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassal primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XJLOR"
/note="Organ: tassels; Vector: HybridZAP; Site:1: EcoRI;
Site:2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 Kb with a 1 kb average."

```

```

FEATURES
source
BASE COUNT 58 a 77 c 97 g 52 t 2 others
ORIGIN
Query Match 59.6%; Score 229.6; DB 35; Length 286;
Best Local Similarity 93.5%; Pred. No. 5,4e-52;
Matches 261; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

```

```

Qy 6 ggcaccaagcctgacatcattgggttgccacggctccagctgagatcgagggaat 65
Db 8 GGCAACAAGCCTGACCTCATTTGTCATGGGCACCGCTCCGAGCTGGAGATCGCGGCCAAG 67
Qy 66 ggcgcacagctgaggaaggagggaag--acggtccgc-gtcctctgttctcgtc 122
Db 68 GCCCGCCACGAGCTGAGGAAGAGGGGCAANGAACGGTCCGCTCGTCTCGTTCGTCTCC 127
Qy 123 tgggaactcttctgatgagcagctcgatgagatcaagagagagcgtcctccctgcgcagtc 182
Db 128 TGGGAACCTCTTGTATGAGCAGTCCGATGAGTACAAGAGAGAGCGTCTCCCTGTGCGGTC 187
Qy 183 acagcaggatcagcatcgagcgggtccactctcgtctggtggcagagtagctcgagcc 242
Db 188 ACAGCGAGGATCAGCATTTGAGGCTGGTCCACCTCTTGGCTGGCAGAGTACGTCTGGAGCC 247
Qy 243 caaggaagcgaattggcatcgacagtgctggcgagct 281
Db 248 CAGGGCAAGGCCATTGGCATCGACAAGTTCCGCGCGAGT 286

```

```

RESULT 15
BE418437/c
LOCUS BE418437 747 bp mRNA EST 24-JUL-2000
DEFINITION SCL033.C06R990602 ITEC SCL Wheat Leaf Library Triticum aestivum
ACCESSION BE418437
VERSION BE418437
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum

REFERENCE 1
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,
S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
Langridge,P., Lazo,G.R., Lin,J., McGuire,P., Oghara,Y.,
Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
Sorrells,M., Warburton,M. and Wenzel,G.
International Triticaceae EST Cooperative (ITEC): Production of
Expressed Sequence Tags for Species of the Triticaceae
Unpublished (2000)
Contact: Cloutier S
Cereal Research Centre, Agriculture & Agri-Foods Canada
Winnipeg MT CANADA
Tel: 204 983 2340
Fax: 204 983 4604
Email: scloutier@em.agr.ca
International Triticaceae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. .747
/organism="Triticum aestivum"
/cultivar="Hatchler Lrl"
/db_xref="taxon:4565"
/clone_lib="ITEC SCL Wheat Leaf Library"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/note="Vector: Lambda ZAP; 1.0 Kbp average insert size."

```

```

FEATURES
source
BASE COUNT 170 a 207 c 177 g 168 t 25 others
ORIGIN
Query Match 58.9%; Score 226.6; DB 35; Length 747;
Best Local Similarity 78.4%; Pred. No. 4e-51;
Matches 279; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

```

```

Qy 5 cggcaccagcctgacatcattgggttgccacggctccgagctggagctcgaggca 64
Db 661 CCGNCACAAGCCTGANTCATCATATAATGAGCNCCTGGTTNTNAGGTAGAGATTGCCGTGAA 602

```

Sat Nov 4 18:11:32 2000

```
Qy 65 tgcggccgacgagctgaggaagaggaggaagacggtccgcgctcgtctcgtctcctc 124
Db 601 GNCCTGCCGAGAGTTGACGAAGAGGGGAAAGCCCGTCCGTTGTGTGTCATTGTTGCTG 542
Qy 125 ggaactctttgatgagcagtcgagatgagtagacaaggaagcgtcctccctgcccacgtcac 184
Db 541 GGAACCTCTTCGACGAGCAGTCAGATGAG-ACAAGGAGAGCGTGTCTCCCTGAGGCCCGTCAC 483
Qy 185 agcggatcagcagcagcgggtccactctcggctggcagaagtagctcggagccca 244
Db 482 CGCAAGGATCAGCATTGAGGCCGGGTCTACCTCGGATGGCAAAAGTACCTCGGGTCCAA 423
Qy 245 agcaaggccattggcatcgacaaagttcgcgcgagtgctcctcgtccgggacgactacaa 304
Db 422 GGGCAAGACCATCGGCATCGACAAGTTGCGCGCCAGCGCACCTGCCCCGAGATCTACAA 363
Qy 305 ggagtacggcatcacccgtggagagcatcattgcaactgcccaagagcttttaagagc 360
Db 362 GGAGTACGGCATCACCGCGGAGAACGTCTATCGCCGACGCCAAGAGCCTGTAAAGATC 307
```

Search completed: November 4, 2000, 11:52:19
Job time: 18256 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:36:23 ; Search time 4075.18 Seconds
(without alignments)
293.660 Million cell updates/sec

Title: US-09-300-482-311

Perfect score: 274

Sequence: 1 ctcatattcttctccacc.....gccaaagcatgccgtcgaccg 274

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*

2: gb_ba2.*

3: gb_om.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl1.*

8: gb_pl2.*

9: gb_pr1.*

10: gb_pr2.*

11: gb_pr3.*

12: em_fun.*

13: em_hum1.*

14: em_hum2.*

15: em_in.*

16: em_om.*

17: em_or.*

18: em_ov.*

19: em_pat.*

20: em_ph.*

21: em_pl.*

22: em_ro.*

23: em_sts.*

24: em_sy.*

25: em_un.*

26: em_vi.*

27: gb_htg1.*

28: gb_htg2.*

29: gb_in1.*

30: gb_in2.*

31: em_ba1.*

32: em_ba2.*

33: em_hum3.*

34: em_hum4.*

35: gb_pr4.*

36: gb_htg3.*

37: gb_htg4.*

38: gb_htg5.*

39: gb_htg6.*

40: gb_htg7.*

41: em_htg1.*

42: em_htg2.*

43: em_htg3.*

44: em_hum5.*
45: gb_pl3.*
46: gb_pr5.*
47: gb_htg8.*
48: gb_htg9.*
49: gb_htg10.*
50: gb_htg11.*
51: gb_htg12.*
52: gb_htg13.*
53: gb_htg14.*
54: gb_in3.*
55: gb_htg15.*
56: gb_htg16.*
57: gb_htg17.*
58: em_htg4.*
59: em_htg5.*
60: em_htg6.*
61: em_htg7.*
62: em_hum6.*
63: gb_htg18.*
64: gb_htg19.*
65: gb_ba3.*
66: em_htg8.*
67: em_htg9.*
68: em_htg10.*
69: em_htg11.*
70: em_htg12.*
71: em_htg13.*
72: em_htg14.*
73: em_htg15.*
74: em_htg16.*
75: em_htg17.*
76: em_htg18.*
77: em_htg19.*
78: em_htg20.*
79: em_htg21.*
80: em_htg22.*
81: em_htg23.*
82: gb_pr6.*
83: gb_pr7.*
84: gb_htg20.*
85: gb_htg21.*
86: gb_htg22.*
87: gb_htg23.*
88: gb_ro.*
89: gb_sts1.*
90: gb_sts2.*
91: gb_sy.*
92: gb_un.*
93: gb_vil.*
94: gb_vil2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	% Match	Query Length	DB ID	Description
c	1	81.4	29.7	6653	8	ATAC016972
	2	81.4	29.7	8497	7	AC006200
	3	68	24.8	1118	5	A93589
	4	62	22.6	1084	7	AF154658
c	5	54	19.7	93234	8	ATAC009465
	6	54	19.7	95310	8	ATAC011437
c	7	47.4	17.3	37898	2	SCE34
	8	46.6	17.0	36224	2	SCD78
	9	46.4	16.9	1288	5	AR007563
c	10	46.4	16.9	161266	8	AP001278
	11	45	16.4	10834	1	AE004020
	12	44.8	16.4	3875	2	RRPHAS
						AC016972 Arabidops
						AC006200 Arabidops
						A93589 Sequence 1
						AF154658 Nicotiana
						AC009465 Arabidops
						AC011437 Arabidops
						AL353862 Streptomy
						AR007563 Sequence
						AP001278 Oryza sat
						AE004020 Xylella f
						X66407 R.ruber ORF

13 44.6 16.3 40790 2 MTCY493 Z95844 Mycobacteri
14 44.6 16.3 77534 1 AF235504 AF235504 Streptomy
15 44.6 16.3 182756 7 AC007789 AC007789 Oryza sat
16 44.4 16.2 43632 2 SC6A5 AL049485 Streptomy
17 44.2 16.1 1593 1 AF113605 AF113605 Streptomy
18 44 16.1 2647 45 OSHPRGP X61280 O.sativa ge
19 44 16.1 28732 1 AF082100 AF082100 Streptomy
20 43.8 16.0 273 2 MTU013938 U13938 Mycobacteri
21 43.6 15.9 154561 8 AP002819 AP002819 Oryza sat
22 43.2 15.8 1230 45 AJ242540 Volvox ca
23 43.2 15.8 19830 2 SC3F9 AL023862 Streptomy
24 43 15.7 701 5 A64741 A64741 Sequence 1
25 43 15.7 38734 2 SC10A5 AL021529 Streptomy
26 43 15.7 68848 2 MTV043 AL022004 Mycobacte
27 42.8 15.6 32274 2 SCC8A AL356892 Streptomy
28 42.8 15.6 133185 49 AC023240 AC023240 Oryza sat
29 42.6 15.5 7836 1 AF047717 AF047717 Streptomy
30 42.4 15.5 989 7 AB027423 AB027423 Oryza sat
31 42.4 15.5 996 7 AB027422 AB027422 Oryza sat
32 42.4 15.5 1167 8 D55714 D55714 Oryza sativ
33 42.4 15.5 1416 7 AB027421 AB027421 Oryza sat
34 42.4 15.5 14472 1 AE003928 AE003928 Xylella f
35 42.4 15.5 18959 1 AE004032 AE004032 Xylella f
36 42.4 15.5 20956 1 AE004082 AE004082 Marine pr
37 42.2 15.4 324 1 AF046836 AF046836 Marine pr
38 42.2 15.4 1187 5 AF046840 AF046840 Marine pr
39 42.2 15.4 1187 5 AF046840 AF046840 Marine pr
40 42 15.3 771 5 A96762 A96762 Sequence 79
41 42 15.3 820 53 AC059012 AC059012 Giardia 1
42 42 15.3 2157 2 PSEPTA D50642 Pseudomonas
43 42 15.3 9344 2 D86947 D86947 Pseudomonas
44 41.8 15.3 833 3 A37833 A37833 Sequence 3
45 41.8 15.3 833 5 AR069871 AR069871 Sequence

13 44.6 16.3 40790 2 MTCY493 Z95844 Mycobacteri
14 44.6 16.3 77534 1 AF235504 AF235504 Streptomy
15 44.6 16.3 182756 7 AC007789 AC007789 Oryza sat
16 44.4 16.2 43632 2 SC6A5 AL049485 Streptomy
17 44.2 16.1 1593 1 AF113605 AF113605 Streptomy
18 44 16.1 2647 45 OSHPRGP X61280 O.sativa ge
19 44 16.1 28732 1 AF082100 AF082100 Streptomy
20 43.8 16.0 273 2 MTU013938 U13938 Mycobacteri
21 43.6 15.9 154561 8 AP002819 AP002819 Oryza sat
22 43.2 15.8 1230 45 AJ242540 Volvox ca
23 43.2 15.8 19830 2 SC3F9 AL023862 Streptomy
24 43 15.7 701 5 A64741 A64741 Sequence 1
25 43 15.7 38734 2 SC10A5 AL021529 Streptomy
26 43 15.7 68848 2 MTV043 AL022004 Mycobacte
27 42.8 15.6 32274 2 SCC8A AL356892 Streptomy
28 42.8 15.6 133185 49 AC023240 AC023240 Oryza sat
29 42.6 15.5 7836 1 AF047717 AF047717 Streptomy
30 42.4 15.5 989 7 AB027423 AB027423 Oryza sat
31 42.4 15.5 996 7 AB027422 AB027422 Oryza sat
32 42.4 15.5 1167 8 D55714 D55714 Oryza sativ
33 42.4 15.5 1416 7 AB027421 AB027421 Oryza sat
34 42.4 15.5 14472 1 AE003928 AE003928 Xylella f
35 42.4 15.5 18959 1 AE004032 AE004032 Xylella f
36 42.4 15.5 20956 1 AE004082 AE004082 Marine pr
37 42.2 15.4 324 1 AF046836 AF046836 Marine pr
38 42.2 15.4 1187 5 AF046840 AF046840 Marine pr
39 42.2 15.4 1187 5 AF046840 AF046840 Marine pr
40 42 15.3 771 5 A96762 A96762 Sequence 79
41 42 15.3 820 53 AC059012 AC059012 Giardia 1
42 42 15.3 2157 2 PSEPTA D50642 Pseudomonas
43 42 15.3 9344 2 D86947 D86947 Pseudomonas
44 41.8 15.3 833 3 A37833 A37833 Sequence 3
45 41.8 15.3 833 5 AR069871 AR069871 Sequence

ALIGNMENTS

RESULT 1
ATAC016972
LOCUS ATAC016972 66653 bp DNA PLN 18-JAN-2000
DEFINITION Arabidopsis thaliana chromosome I BAC F23N20 genomic sequence,
complete sequence.
AC016972
AC016972.5 GI:6714311
HTG.
thale cress.
Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 66653)
Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B., Wu,D.,
Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R., Barnstead,M.E.,
Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome I BAC F23N20 genomic sequence
Unpublished
2 (bases 1 to 66653)
Lin,X. and Kaul,S.
Direct Submission
Submitted (09-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 66653)
Lin,X.
Direct Submission
Submitted (18-JAN-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Jan 18, 2000 this sequence version replaced gi:6630537.
Address all correspondence to:at@tigr.org

BAC clone F23N20 is from Arabidopsis chromosome I and is near the
molecular marker m453.
The orientation of the sequence is from SP6 to T7 end of the BAC

Genes were identified by a combination of three methods: Gene
prediction programs including GAIL (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://www.cbs.dtu.dk/net/gene/cbsnetgene.html), and NetPlantGene
(http://www.cbs.dtu.dk/net/gene/cbsnetgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tldb/at/at.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers
1..66653
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="I"
/map="m453"
/clone="F23N20"
1..17670
/note="overlap with BAC clone F15H11
(AC008148:92846..110514)."
1..15580
/note="the annotation for genes within this region can be
found in the overlapping bac F15H11 sequence 92846-108425"
1
/note="37813 nt before this point were not included in the
submitted sequence, due to overlap with another BAC
(F3117)"
/rpt_family="TAAAAA)n"
complement(1088..1149)
/rpt_family="POLY_A"
complement(3778..3842)
/rpt_family="(GAAAA)n"
complement(3778..3821)
/rpt_family="POLY_A"
complement(4015..5903)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."
complement(4183..4303)
/note="ATREP3|ATREP3 a consensus."
complement(4242..4477)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."
complement(4245..4396)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."
complement(4404..4475)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."
complement(4461..4600)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."
complement(4526..4597)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."
complement(4613..4670)
/note="ATREP5|ATREP5 An AT-rich repetitive sequence - a
consensus."
complement(4662..4832)
/note="ATREP4|ATREP4 An AT-rich repetitive sequence - a
consensus, from T23015 64288 to 65702 1415 nt, also
present in intron 8 of Arabidopsis thaliana gene for

overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/tldb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://arthur.epm.ornl.gov/pub/xgrail>), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/GENSCAN.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are annotated as tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSLH/WashU/ABI consortium for sequencing BAC clones F623, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES	Location/Qualifiers
source	1..84957 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="II" <1..2020
misc_feature	/note="Sequence from clone F23H14" complement(join(<463..924,1053..1163,1265..1417,1513..1572,1662..1967,2287..2435,2515..2584,2718..2921,3271..3465,3673..3945,4026..4187,4325..4455,4541..4700,4817..5005,5145..>5219)) /gene="At2g01130" complement(<463..>5219) /gene="At2g01130" /codon_start=1 /note="F10A8.1" complement(join(<463..924,1053..1163,1265..1417,1513..1572,1662..1967,2287..2435,2515..2584,2718..2921,3271..3465,3673..3945,4026..4187,4325..4455,4541..4700,4817..5005,5145..5219)) /gene="At2g01130"
gene	/product="putative ATP-dependent RNA helicase A" /protein_id="AADI4515.2" /db_xref="GI:6598524"
CDS	/translation="MLDSRTSLPAFKQKHVSALTQISQVIVISGTCGKGTQIPQF ILLESFANRGAFSIICTOPRISAMSVSERVAGEQLGESGVKRLGVKGRD TRLEPCTGILLRLLDVRLNRYGTHVVDIEIHERGMNEDFLIILKDLSSRELKL ILMSATDAELFSSTFGGAGVIYIPGTFYVRSHFLEDELEMTYRLTPYNOIDDDYQ ERTWMKQIIPKRSQITFDALRAADKFESPETRESLSCTWTPDCIGNLFIEFLIC NICNEGPGILIFLTGQWDDISLKEQLIHPHIFGNPLVMLLACHGSMFEQRLIF EPPASGRVIRVATIAETSIINDVAEIVDCRAKETSYDALNTPCLPLSPSKYS AQORGRAGRVPGQCYHLYPKVCYDAEYQLPEILTPHLSCLQTLKSLGISE FLSRALQSPLELAVOKATAFLKIIGALDENDLTLGRLYLSKPEKRLGMLILGAI LGCLDPLITVAGLSVRDPLTPDCKDLAEAAKQSFSDHSDHLAVRAVEGWKKA EESAVDYCNWNLFSIQSLRALDSLRKEFFSLIKDITGLIDGNPSTCNSEGNANLTRA

misc_feature	VICVMPGICSVVHNSRSLKTMEDQVLLYSNSNARETKIPYPWLVFNKIKVN SVFLRDSACSDSTLIILFGGSIKGDITDGLHMLGGLYFEMKPDVAEIIYQTLKKELD ELIONKLLNPKYQMAHRELLSALRLLVSEDCDGRVFGHQILRLPLETSALSTKPSL FSRTSPGPDGNSKSLQTLITRAGYIVPMYTKQLKNKFPQTVENETQIMGQPCS NKSAAEKAAAEAIQWLKGGKESHEQVHMSKLLKGGKKDHL" 2021..>84957 /note="Sequence from clone F10A8" complement(join(6441..7261,7365..7474,7570..7839, 7952..>8107)) /gene="At2g01140" complement(6441..>8107) /gene="At2g01140" /note="F10A8.2" 6543..6582 /rpt_family="CAAAA)n" complement(join(6622..7261,7365..7474,7570..7839, 7952..8107)) /gene="At2g01140" /codon_start=1 /product="putative fructose-bisphosphate aldolase, plastic form" /protein_id="ADI4543.1" /db_xref="GI:4262250" /translation="MASASFVKPNTLSPTWIGQSFATHTSSSPPRVSFAIRAGAY SDELVKTAISAPGRLAIDESNATCGKRLAIGLDNEDNQAIRQLLITPGLG DYISGILFEETLYQSTKDTQFVDCLRDANIVFGIKVKGSLAGSNEESWCQGLD GLASAEYKQGARFAKWTVPVPGPSALAEKAAGLARYAAISQDNGLVPIVE PEILLGDHPITERTLEVAEKVWSEVFFLAQNVMFEGILKPSMTVPGAHKNKASP ETVADFTLMLKRPVAVPGIMFELSGSGSAEATNLNNAKQSPNPHVVSFSVARAL QNSVLTWQGAPEKIEASQALLVRKANSLAQLGKYSAGEGENEDAKKGMFVKYTY" complement(9948..10066) /rpt_family="TAAA)n" <12241..12943 /gene="At2g01150" 12241..12943 /gene="At2g01150" /note="F10A8.3; identical to GP:AF078823" 12241..12762 /gene="At2g01150" /codon_start=1 /product="RING-H2 zinc finger protein (RHA2b)" /protein_id="AADI4516.1" /db_xref="GI:4262223" /translation="MRCKVVAARTTRTKTKTISKTMGLQGLSDVSDSTPLM LIALLATFRRVRSLLLPSSAPVVVTSNLSLADNLNRLFSYRSDNAASDCIV CLSKLTGEVRRKLCRHVFHKKQCLEGLQHLNFCPLRSLPLPHHGHGSDASIS APPLKSTSPASSH" 12263..12319 /rpt_family="CAA)n" 13681..13754 /gene="At2g01160" /product="tRNA-Ala" /anticodon="(pos:13715..13717,aa:Ala) 13681..13754 /note="At2g01160" /gene="F10A8.4" complement(join(<13980..14327,14404..14533,14615..14904, 14978..15159,15239..15480,15649..15909,15978..16109)) /gene="At2g01170" complement(<13980..16109) /note="At2g01170" /gene="F10A8.5; identical to GP:AF019637" complement(join(13980..14327,14404..14533,14615..14904, 14978..15159,15239..15480,15649..15909,15978..16078)) /gene="At2g01170" /codon_start=1 /product="putative amino acid permease" /protein_id="AADI4517.1" /db_xref="GI:4262224" /translation="MGLGGQSFVPMDSQVRLKELGYKQELKRLSVFSNFAISFS IISVLTITTYNTGTRFGGTFLVYGFAGSFTMCVGLSNAEICSSTPSGGLYYK SAMLGRWAPLASMWTGKVVYQLQWAVTASDFSLAQIQLVILISTGGNGGYYK GSDVYVIGIHGILTFHALLINSIPVLSVLSFIQALAAWNLGLVLMILLIPVSTERA
repeat_region	
CDS	
repeat_region	
tRNA	
gene	
mRNA	
gene	
CDS	

```

TTKFTNFTNDNGLITSYAYIFVLGLMSQYITITGYDASAHWTETVDADKNGPRG
IIISAIGSILFCWGYIIGISTAVTDIPSLISETNNSGYIAEIFYLAFKRNFGSGTG
GIVCLGVAVAVAFCCGSSVTSNMAIYAFSRDGAEMPSPLWHKVNREVPINAVWLS
ALISFCNALVSLGSIYAFQAMVSIATIGLYIAYAIPIILRVTLARNTFVCPGPSLGRY
GVGVWAVLWVVTISVLFSLPVAIPITAEITLNTPTVAVAGLVAILTSLYWLFSARHWF
TGPISNLS"
complement(<18798..>19706)
/gene="At2g01180"
/gene="At2g01180"
/complement(<18798..>19706)
/gene="At2g01180"
/complement(<18798..>19706)
/gene="At2g01180"
/codon_start=1
/product="putative phosphatidic acid phosphatase"
/protein_id="AADI4518.1"
/db_xref="GI:4262225"
/translaton="MOEIDLVSHTIKSHGRVASKKHHDWIIILVILIAIEIGLNLISL
FYRVGKDMTDLKYPFKDNTPIWSPVYAVILLPIIVFCVFLKRTCVVDLHHSILG
LLFAVLITGVITDSIKVATGRPRPNFWRFCFDPGKELYDALGVCHGKAAEVKEGKH
SFGSGHTWSFAGLTFLSLVLSGKIKAFNNEGHVAKLCLVIFPLLAACLVGISRVDDY
WHHQDVFAGALLIGTLVAAPCYRQFPNYPNYPHEEGWPYAFYKAAQERGVPTVSSQNGD

Query Match
Best Local Similarity 29.7%; Score 81.4; DB 7; Length 84957;
Matches 121; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 88 accccatttcacccagagagccgcatgagcgcgcgcctcctccaccctcct 147
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61589 ATCTCTCTTCATTCATCGGACAAATCTTTGCCGCTTTGATGTTCACCGC 61530

Qy 148 cccctccgctcctccaccagacgatttgaagaaatcgccgctcacaagccgctcg 207
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61529 CTCAGCCCATGAATTTAACACAGACGAGCTCAACGCTATCGCGCTTCAAAAGCGGTG 61470

Qy 208 agtacggagtcggcatgctcgcctcgcctgagcgcgcctcaccgccaagcatgccg 267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61469 AATTCGTCGAGTCGGCATGTTCTCGGTCTCGGAACCGCTCCACCGCAACACGCGC 61410

Qy 268 tcgaccy 274
|||||

Db 61409 TCGACCG 61403

RESULT 3
A93589
LOCUS 1118 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent WO9737028.
ACCESSION A93589
VERSION A93589.1 GI:6741794
KEYWORDS
SOURCE
ORGANISM Spinacia oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Caryophyllales; Magnoliopsida; Chenopodiales; Spinacia.
1 (bases 1 to 1118)
Scharrenberger, C. and Lerchl, J.
RIBOSE-5-PHOSPHATE ISOMERASE (D-RIBOSE-5-PHOSPHATE KETOL ISOMERASE,
EC 5.3.1.6)
Patent: WO 9737028-A 09-OCT-1997;
SCHNARRENBARGER CLAUS (DE); BASF AG (DE)
FEATURES
Location/Qualifiers
source
1..1118
/organism="Spinacia oleracea"
/db_xref="taxon:3562"
1..24
25..894
/note="unnamed protein product"
/codon_start=1
/db_protein_id="CAB69583.1"
/db_xref="GI:6741795"
/translaton="MASAAFLSLPSTSTFNHATTKLLNLKFLHNHRNKPFTTTI

mRNA
gene
CDS
5' UTR
CDS

```

```

KSLSSPSPTPVLTOQDDKLAAEKAVDSVKSQWVLGLGTGTAFAVSRIGELLSAGK
LNIIVGIPTSKRTAEQAASGLKPLSLVDDHPRIDLAIDGADVEDPDLNLVKGKRGALL
RKMWBAASDKFIIVVDDTKLVGLGSLAMPVEVQFCWKYINLRLQEIIFKELGCE
AKLRMGSSSPYTDNSNYIVDLIYPTSTIKDAAGREISALEGVVHGLFLGMASEV
IIAGKTGVSVTK"
895..1118
BASE COUNT 261 a 264 c 258 g 335 t
ORIGIN

Query Match
Best Local Similarity 24.8%; Score 68; DB 5; Length 1118;
Matches 95; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 135 ctcaccctctccctccctccctccctccctccctccctccctccctccctccctcc 194
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 222

Qy 195 tacaagggcgcgtgagtcgtgagtcgtgagtcgtgagtcgtgagtcgtgagtcgtg 254
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 GAAAAAGCCGTCGACTCCGCAATCCGCGCATGGTTCTCGGTCTCGGAACCGGAAGTACT 282

Qy 255 gccaaagcatgcccgtgacgcg 274
|||||

Db 283 GCCGATTTGCTGCTCTCGCG 302

RESULT 4
AF154658
LOCUS 1084 bp mRNA PLN 15-JUN-2000
DEFINITION Nicotiana tabacum clone PR49 mRNA sequence.
ACCESSION AF154658
VERSION AF154658.1 GI:8547158
KEYWORDS
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
1; Solanales; Solanaceae; Nicotiana.
Sheehan, M.J., Wang, J., and Timko, M.P.
Characterization of cDNAs Differentially Expressed in Roots of
Tobacco (Nicotiana tabacum cv Burley 21) During the Early Stages of
Alkaloid Biosynthesis
Unpublished
2 (bases 1 to 1084)
Sheehan, M.J., Wang, J., and Timko, M.P.
Direct Submission
Submitted (27-MAY-1999) Biology, University of Virginia, 044 Gilmer
Hall, Charlottesville, Virginia 22903, USA
FEATURES
Location/Qualifiers
source
1..1084
/organism="Nicotiana tabacum"
/cultivar="Burley 21"
/db_xref="taxon:4097"
/clone="PR49"
/tissue_type="root"
/note="similar to ribose 5-phosphate isomerase"
286 a 245 c 250 g 302 t 1 others

BASE COUNT 286 a 245 c 250 g 302 t 1 others
ORIGIN

Query Match
Best Local Similarity 22.6%; Score 62; DB 7; Length 1084;
Matches 74; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 163 tcaccagacgatttgaagaaatccgcctcacaagcgcgtcagtcagtcagtcagtcag 222
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 TAACCCAGAGCACTCAAAAACCTCCGCCGCGGATAAGCGGTCGTAATACGTGAAAACGC 307

Qy 223 gcattgctcgcgtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 256
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 GAATGGTCTTGGTCTAGGCACCTGGTTCAACCGC 341

```

```

RESULT 5
LOCUS ATAC009465/c
DEFINITION ATAC009465 93234 bp DNA PLN 29-SEP-1999
           Arabidopsis thaliana chromosome III BAC T9J14 genomic sequence,
           complete sequence.
ACCESSION AC009465
VERSION   AC009465.5 GI:5932531
KEYWORDS  HTG.
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
           Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
           Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
           Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 93234)
AUTHORS   Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,
           Rowan, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,
           Bowman, C.L., White, O., Niernan, W.C. and Fraser, C.M.
           Arabidopsis thaliana chromosome III BAC T9J14 genomic sequence
           Unpublished
           2 (bases 1 to 93234)
           Lin, X. and Kaul, S.
           Direct Submission.
           Submitted (24-AUG-1999) The Institute for Genomic Research, 9712
           Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org
           3 (bases 1 to 93234)
           Lin, X.
           Direct Submission
           Submitted (29-SEP-1999) The Institute for Genomic Research, 9712
           Medical Center Dr., Rockville, MD 20850, USA
           On Sep 29, 1999 this sequence version replaced gi:5809709.
           Address all correspondence to:
           Xiaoying Lin
           The Institute for Genomic Research
           9712 Medical Center Dr.
           Rockville, MD 20850, USA
           e-mail: xlin@tigr.org
           BAC clone T9J14 is from Arabidopsis chromosome III and is near the
           molecular marker ml172.
           The orientation of the sequence is from SP6 to T7 end of the BAC
           clone.
           Genes were identified by a combination of three methods: Gene
           prediction programs including GRAIL (available by anonymous ftp
           from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
           Washington), GenScan (Chris Burge,
           http://gnomic.stanford.edu/~chris/genSCANW.html), and NetPlantGene
           (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
           complete sequence against a peptide database and the Arabidopsis
           EST database at TIGR (http://www.tigr.org/tldb/at/at.html).
           Annotated genes are named to indicate the level of evidence for
           their annotation. Genes with similarity to other proteins are named
           after the database hits. Genes without significant peptide
           similarity but with EST similarity are named as 'unknown' proteins.
           Genes without protein or EST similarity, that are predicted by more
           than two gene prediction programs over most of their length are
           annotated as 'hypothetical' proteins. Genes encoding tRNAs are
           predicted by tRNAscan-SE (Sean Eddy,
           http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
           identified by RepeatMasker (Arian Smit,
           http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
           genomic sequence that are not annotated as genes but have predicted
           exons by GRAIL are annotated as misc features.
FEATURES             Location/Qualifiers
     source            1..93234
                       /organism="Arabidopsis thaliana"
                       /cultivar="Columbia"
                       /db_xref="taxon:3702"
                       /chromosome="III"
                       /map="mil172"
                       /clone="T9J14"
                       /complement(l..945)
     misc_feature      /note="overlap with BAC clone T12H1 (AC009177:1..945)."
misc_feature          3576..3617
                       /note="exon predicted by xgrail, quality marginal"
                       complement(4024..4148)
                       /note="exon predicted by xgrail, quality
                       marginal_shadowexon"
                       4493..4537
                       /note="exon predicted by xgrail, quality marginal"
                       4806..4964
                       /note="exon predicted by xgrail, quality marginal"
                       complement(5275..5322)
                       /note="exon predicted by xgrail, quality
                       marginal_shadowexon"
                       5360..5466
                       /note="exon predicted by xgrail, quality good"
                       5516..5613
                       /note="exon predicted by xgrail, quality excellent"
                       complement(6088..6145)
                       /note="exon predicted by xgrail, quality
                       marginal_shadowexon"
                       6651..6871
                       /note="exon predicted by xgrail, quality good"
                       complement(6987..7047)
                       /rpt_family="(TAA)n"
                       complement(7253..7293)
                       /note="exon predicted by xgrail, quality
                       marginal_shadowexon"
                       7364..7412
                       /note="exon predicted by xgrail, quality good"
                       7681..7820
                       /note="exon predicted by xgrail, quality good"
                       7979..8097
                       /note="exon predicted by xgrail, quality excellent"
                       join(-8161..8331,8463..8581,9011..9108,9205..9264,
                       9340..9386,9475..9669,9764..9810,9892..9958,10061..10161,
                       10253..10354,10446..10584,10669..10731,10832..10915,
                       11142..11489)
                       /gene="T9J14.2"
     gene              <8161..>11489
                       /note="similar to sodium proton exchanger Nhxl GB:AAD16946
                       [Arabidopsis thaliana]"
                       join(8161..8331,8463..8581,9011..9108,9205..9264,
                       9340..9386,9475..9669,9764..9810,9892..9958,10061..10161,
                       10253..10354,10446..10584,10669..10731,10832..10915,
                       11142..11489)
                       /gene="T9J14.2"
                       /codon_start=1
     CDS                /product="putative sodium proton exchanger"
                       /protein_id="AAD56988.1"
                       /db_xref="GI:5932533"

```


initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid D78 lies to the right of Q11 on the AseI-D genomic restriction fragment.
 Location/Qualifiers

FEATURES

source

1. 36224

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid D78"

1. 817

/gene="SCD78.01"

<1. 817

/gene="SCD78.01"

/note="SCD78.01, ftsK /spoIIIE family protein, partial CDS within putative integrated plasmid, len: 271 aa; similar to many members of the ftsK /spoIIIE family e.g. SPI_STRAZ (EMBL:S58719) sporulation-regulatory protein from Streptomyces azureus plasmid pSA1.1 (303 aa), fasta scores: opt: 736 z-score: 1014.5 E(): 0, 49.0% identity in 247 aa overlap and TRSA_STRAM (EMBL:Z19593) transfer protein trasa from Streptomyces ambofaciens plasmid pSAM2 (306 aa), fasta scores: opt: 222 z-score: 676.5 E(): 2e-30, 47.8% identity in 203 aa overlap. Also similar to SC6A9.34, a protein within another putative integrated plasmid (on Streptomyces coelicolor cosmid St6A9)(E): 5.4e-24, 45.7% identity in 247 aa overlap). Contains PS00017 ATP /GTP-binding site motif A (p-loop).
 /codon_start=2

/transl_table=11

/product="ftsK /spoIIIE family protein"

/protein_id="CAA22206.1"

/db_xref="GI:3928710"

/translation="IGALETGGAVNNLRLVPHWLIAGTRSGKSTLLARVITQAPQ PVALVGTDCRGMEGLFADRLSALATSRREAVLTAIVDQIERSACRGVPSV WELPDKRPVGVVVLVDEIAELVSDGTQSKSEAEQCSFTLLIRLAQLGAALGLHLV AGORVSDLGPGVTAARLAQGLGRICHVRNDPGRMETGLDNKDAVAQALTAQERG VAVCTGPDGWARASHLPTDEAVATARKSMTPELPALDAUVALEGGCK"
 1. 4443

misc_feature

/note="possible integrated element"

71. 34

/gene="SCD78.01"

/notes="PS00017 ATP /GTP-binding site motif A (p-loop)"

814. 1011

/gene="SCD78.02"

814. 1011

/gene="SCD78.02"

/note="SCD78.02, small hydrophobic protein, len: 65 aa;

within putative integrated plasmid"

/codon_start=1

/transl_table=11

/product="hypothetical protein SCD78.02"

/protein_id="CAA22207.1"

/db_xref="GI:3928711"

/translation="MISEGTAFGFAVVGIIITVLLVRSRDVRGWEAVCVGLFGLVGG TPVLEFIHGLVTVWISGFSHT"

1308. 2666

/gene="SCD78.03"

1308. 2666

/partial

/gene="SCD78.03"

/note="SCD78.03, possible plasmid replication initiator protein within putative integrated plasmid, len: 452 aa; similar to REPS_STRAM (EMBL:Z19594) replication initiator protein from Streptomyces ambofaciens plasmid pSAM2 (459 aa), fasta scores: opt: 712 z-score: 983.7 E(): 0, (43.7% identity in 467 aa overlap). Note that there is no possible initiation codon upstream of the start of homology (or close downstream). The CDS given here has an arbitrary start"

/transl_table=11

/product="putative thiamin biosynthesis protein thic"

/protein_id="CAA22210.1"

/db_xref="GI:3928714"

/translation="SMKISQDIRRHQSGSKGEIEGMAQSKFEFAAGNRYLPIAD"

4655. 6103

/gene="SCD78.06"

4655. 6103

/gene="SCD78.06"

/note="SCD78.06, unknown, len: 482 aa; similar to

17.0%; Score 46.6; DB 2; Length 36224;

Best Local Similarity 52.9%; Pred. No. 9;

Matches 100; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

/codon_start=1
 /transl_table=11
 /product="putative plasmid replication initiator protein"
 /protein_id="CAA22208.1"
 /db_xref="GI:3928712"

/translation="MRHLPEADRALRIQAQDPDFPRWLOQITATGSCAHPVHLSGHTT TLDNATGEILRHVDTRNEPGERLLVRCNRRSTVCPCSLRHAGDTFLVLRAGLLGGK NVPASVTRPRLFTVLTAPSGAVHVRAGOTGPRRDGACGHRPLGCGCTVHAPDAPA VGQPLCPDCYDYTAHVLAHASKLWDREVIDVRRELASSAGIVSRFAHARLSFAR IAEYKRAAVHVHTVYVRLDGPDPDEPAWGTDRLRTRAVHASQRVTVRTPYSLAT GELELSWGVTQDVCVYLHAGDGPDDDAVAYVAKYVSKGTDEGAGTDRKVTTCDDID SARVSRHVTRLMTHTCRLGLGPEYLRRLTWTHTLGRGHILTKRSRSTTYTAALRA QRAHHGHTDTPDAITDAHWRYVSGHTPGAAIIAAGVAEDIAQNAARVADALSAGG AT"

gene

complement(2805..4292)

/gene="SCD78.04c"

complement(2805..4292)

/note="SCD78.04c"

/gene="SCD78.04c, possible integrase/ recombinase within putative integrated plasmid, len: 495 aa; similar to TR:Q37839 (EMBL:D38173) ORF469 protein (possible site-specific recombinase) from Actinophage R4 (469 aa), fasta scores: opt: 261 z-score: 320.1 E(): 1.4e-10, 29.1% identity in 306 aa overlap and TR:O06604 (EMBL:Z95586) MTCY336.18 (RV1586c) possible integrase from M. tuberculosis integrated phage phiRv1 (469 aa), fasta scores: opt: 371 z-score: 270.8 E(): 7.8e-08, (26.6% identity in 489 aa overlap). Probable coiled-coil from 414 to 451 (38 residues) Max score: 1.608 (probability 0.99);
 /codon_start=1

/transl_table=11

/product="putative integrase /recombinase"

/protein_id="CAA22209.1"

/db_xref="GI:3928713"

/translation="MLAADPNLKCVCVYARISDFGRVVDANGIEDQHDMSEAAARRFG WLIVRYTNDKSKASKESVNDPQELLADLAAGATPEGVPVHGVMVNDRLVRRPS DMERLQTSQDGRVYHDSGLQDYAEFEIKGLGVAMSLSETRKKRRSRNSHR SRAIRGQSVAMRPFGEDEKVTLPDPAEAIARTAVHDVIAIASITRKKWKAAGFI TSRGNFQYQTPVKGVLVNRKCGYREIKGEIVRGDDQPIVGWEAIVTPQWFAVTA KIREHGHTGTVRGGLVHKYLLTNILRCGNVLEDTCVNNKMGITKANWMLKYOHAYM CKTKVDGCKNTYKRGDKTKTIEELVIAKLERDAATKAQDVPWDKAEALERALQSR RELRRWHDDTDIDDEAFRNLPVLEERIKELRVQDKAHEAEAEADLRKS WGAKTLTQKREAMKAVLGAVIALPGGKNGKTFDPLLPKPVWKTSE"

4444. 4575

/gene="SCD78.05"

/note="thic"

4444. 4575

/partial

/gene="SCD78.05"

/note="SCD78.05, thic', probable thiamin biosynthesis protein thic, partial CDS, len: 43 aa; similar to the extreme C-terminus of e.g. THIC_ECOLI thiamin biosynthesis protein thic (631 aa), fasta scores: opt: 97 z-score: 137.7 E(): 2, 46.2% identity in 39 aa overlap and THIC_MYCTU (547 aa), fasta scores: opt: 143 z-score: 272.4 E(): 6.4e-08, 53.1% identity in 49 aa overlap. Note that this gene appears to have been interrupted by the integration of the upstream element"

/codon_start=1

/transl_table=11

/product="putative thiamin biosynthesis protein thic"

/protein_id="CAA22210.1"

/db_xref="GI:3928714"

/translation="SMKISQDIRRHQSGSKGEIEGMAQSKFEFAAGNRYLPIAD"

4655. 6103

/gene="SCD78.06"

4655. 6103

/gene="SCD78.06"

/note="SCD78.06, unknown, len: 482 aa; similar to

17.0%; Score 46.6; DB 2; Length 36224;

Best Local Similarity 52.9%; Pred. No. 9;

Matches 100; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Query Match
 Best Local Similarity 52.9%;
 Matches 100; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

REFERENCE
AUTHORS
1 (bases 1 to 10834)
Silvestri,M.L., Siqueira,W.J., de Souza,A.A., de Souza,A.P.,
Terenzi,M.F., Truffi,D., Tsai,S.M., Tshako,M.H., Vallada,H., Van
Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M.,
Meldani,J. and Setubal,J.C.
TITLE
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis, Sao Paulo, Brazil
JOURNAL
MEDLINE
REFERENCE
2365717
2 (bases 1 to 10834)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
Kemper,E.L., Klafajma,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A.A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silva,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
Vettore,A.L., Zago,M.A., Zatz,M., Meldanis,J. and Setubal,J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil
REMARK
FEATURES
source
Location/Qualifiers
1. .10834
/organism="Xylella fastidiosa"
/db_xref="taxon:2371"
122. .775
/gene="XF2008"
122. .775
/gene="XF2008"
/note="similar to GI|4981730 (percent identity: 39 %/query
alignment coverage: 68.7 %/subject alignment coverage:
83.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84810.1"
/db_xref="GI:9107123"
/translation="MRSMLPSPDFTATLISNRSDPVLFFTHQDFTAMSSNTPRV
VFEKQRMIVDGVSEYRHPGGLAAIILAVTPERRILFVQFRIPQIQTIEPQA
GLVGDIDAGESIEQAAIIELEFEETGWTADHILLIGPVSSGASSERIAFVATGLRK
VSSGGGVDEIDVYHEIPRTQAAALVQKMHAGFAMPDKLWALWLDHLDGTPIK"
complement(899. .2239)
/gene="XF2009"
complement(899. .2239)
/note="similar to GI|1653657 (percent identity: 46 %/query
alignment coverage: 98.9 %/subject alignment coverage:
100.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder/Start codon shift: -75"
/codon_start=1
/transl_table=11

/product="aminopeptidase P"
/protein_id="AAF84811.1"
/db_xref="GI:9107124"
/translation="MNRIVKHTGIAAEYCRBRRLMKVGPQGIWVLPAAPEVRVS
RDTHYPRQSDSDFWLCGPEPDAVLVIFGRCQVLLFCRDEPDEANDGPRAGY
DGAIEQYMDSDAYPIDDLDELPLGREGSOYVHFGDFEDLRLIGWNRVDRQVR
QAASPHFEIDLGHLLHEORLSYDIIVLQCAADISVEAHLAALRYAHGVSYYL
YAELEFRADSPAYTSIVAAGANACVILHYRANAECSRDGDIIVTADAGYRGVAA
DITTEPVNGRSPQALYDLGAAVNLADQVGLPYFAGHAAVQVTLTEGLLRL
GLLHGTLEDNLADQSKYKFEYHKTCHWGLDVHDVGDYRIDGSRRLLEPGMVTIEPG
LYLVDDTAVHPKRWGIGRTEDDDVLTEGCHVLTGALPSADEIEREANAAR"
complement(2236. .2793)
/gene="XF2010"
complement(2236. .2793)
/note="similar to GI|140683 (percent identity: 36 %/query
alignment coverage: 91.4 %/subject alignment coverage:
87.1 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84812.1"
/db_xref="GI:9107125"
/translation="MESPMHLPVIAVOQESQMGSLVTAPELHGSLSGLLAGGGNG
PDWLAMILADAEVAAPKPSVLSRQATSLQEDPDFAFQLLADGATLAAADAL
FEWCAFLGFGFLAAHSRSLVSAEGDETLRLAKLAQASVDDFDMNEEDGSLIEIE
EFVRVAVLLHGDCLIGCAPQPLN"
2986. .3207
/gene="XF2011"
2986. .3207
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84813.1"
/db_xref="GI:9107126"
/translation="MDQPDALTELHSLAAQLERLEYCQRLAEENHLRCOOHOMMLE
RELLAKNEQARSVEANVTRIKTLEHT"
3204. .3503
/gene="XF2012"
3204. .3503
/note="similar to SPIP45580 (percent identity: 27 %/query
alignment coverage: 91.9 %/subject alignment coverage:
83.5 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84814.1"
/db_xref="GI:9107127"
/translation="MSDTEPVSIIRILDREYTVGVGIERRKGLIAAQLLDHMRREIRG
SNRMASIDRIAVIALNALNAHELQQLHEKHADEERALLTLDLQORLNAMDYSH"
4005. .4616
/gene="XF2013"
4005. .4616
/note="similar to SPIP44905 (percent identity: 35 %/query
alignment coverage: 88.2 %/subject alignment coverage:
95.7 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84815.1"
/db_xref="GI:9107128"
/translation="MIQTSKPMTDNRNLSIRQHMARRRAIHKHARLAANVTLAEHLIT
LPFAQTGAVAGYWMADGEIPLRWQTLPAQVRYCLPILAGTKLRFPAHPQGPLKQ
NRYGIPDVTIDETLTPESMTLVVTVPLVGFDTQGRHMGGGWDRSFAFRNQSP
PWLGVGVSVQVSTLPTAEWDICVDAICTELTLLPATVST"


```

gene      4637. .5080
/ gene="XF2014"
CDS
/ gene="XF2014"
/ note="similar to GI:6458261 (percent identity: 53 %/query
alignment coverage: 98.6 %/subject alignment coverage:
89.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84816.1"
/db_xref="GI:9107129"
/translation="MKSEPDFAFIDDLQVRGTEPWNGIRNYQARNFIRDKIHGDSVL
FYHNSCKPGVIGLAKIASAYDESDQFNSDIYDPKARQEPWILVDIAFERKLA
RTTLERIKQHALGEGFPLTRGNRLSVFPVTTQWKLLLSLE"
/ gene="XF2015"
/ gene="XF2015"
/ note="similar to SP|P27252 (percent identity: 57 %/query
alignment coverage: 84.5 %/subject alignment coverage:
94.5 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder/Start codon shift: -147"
/codon_start=1
/transl_table=11
/product="ribose-5-phosphate isomerase A"
/protein_id="AAF84817.1"
/db_xref="GI:9107130"
/translation="METATVTRVNSTQPNACAPSHIPIIINPMSEAKRRRAEKAIE
YVNDLIGVGTSVYFIDALGRTPKIKGAVSSSEOSTAHLKHOGTEVLEKHTG
TLPLYVDGADQCPYKRLIKGGASLTREKIIAEASKOFICIIDPNKQVATLQKPLP
IEVTPARSLVARQINARTDQGPVWREGVITDNGNVILDVHHLRIITDPVKLSQELNQI
PGVVCGLPARRCADVVIIGSEPPHIL"
/ gene="XF2016"
/ gene="XF2016"
/ note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84818.1"
/db_xref="GI:9107131"

Query Match      16.4%; Score 45; DB 1; Length 10834;
Best Local Similarity 70.6%; Pred. No. 23;
Matches 60; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 188  cgcgcctacaagccgctgagtcagtcgagtcgcgcgctgctcgcgcctagcaccgg 247
Db 5163  CGCGCGGAAGGCCATCGAATACGTCGGAACGATATGATTCGCGCGTGGCACC GG 5222

QY 248  ctcacgcgcaagcatgcgcctgcag 272
Db 5223  CTCACAGTCGCTACTTCTATCGAC 5247

RESULT 12
RRPHAS
LOCUS      RRPHAS      3875 bp      DNA      BCT      30-NOV-1993
DEFINITION      R.ruber ORF1, phbCrr, ORF3 and ORF4.
ACCESSION      X66407.4
VERSION      X66407.1
KEYWORDS      PHA-synthase.
SOURCE      Rhodococcus ruber.
ORGANISM      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE      1 (bases 1 to 3875)
AUTHORS      Steinbuechel, A.
TITLE      Direct Submission

```

JOURNAL

REFERENCE
AUTHORS
TITLEJOURNAL
REFERENCE
AUTHORS
TITLEJOURNAL
MEDLINE
FEATURES

Source

CDS

terminator
terminator
gene

CDS

CDS

terminator
CDS

```

Submitted (25-MAY-1992) A. Steinbuechel, Inst. fur Mikrobiologie
der Georg-August-Universitaet, Grisebachstr. 8, W-3400 Goettingen,
FRG
2 (bases 1 to 3875)
Pieper U. and Steinbuechel, A.
Identification cloning and sequence analysis of the
poly(3-hydroxyalkanoic acid) synthase gene of the gram-positive
bacterium Rhodococcus ruber
FEMS Microbiol. Lett. 96, 73-80 (1992)
3 (bases 1 to 3875)
Pieper U. and Steinbuechel, A.
Identification, cloning and sequence analysis of the
poly(3-hydroxyalkanoic acid) synthase gene of the gram-positive
bacterium Rhodococcus ruber
FEMS Microbiol. Lett. 75 (1), 73-79 (1992)
92406022
Location/Qualifiers
1. .3875
/organism="Rhodococcus ruber"
/strain="NCIMB40126"
/db_xref="taxon:1830"
83. .598
/ note="ORF1"
/codon_start=1
/transl_table=11
/protein_id="CAA47034.1"
/db_xref="GI:46399"
/db_xref="SPTREMBL:Q53048"
/translation="MRCPLSRTGYSRIIRGDDHPRQSPNAPSARGSGRSRYADP
MEDDASRPPTLGELIRQRELAELPMQLAAWVGISNPIYSQIERNLRAPSERVLOAI
AEQHLSDADALTAEVGRPDGSAVVRAIHEDPDLTNAQRSLVEMYEAFFREVTVGKR
RRGARPDGTE"
595. .638
terminator
644. .698
terminator
774. .2462
/ gene="phbCrr"
774. .2462
/ gene="phbCrr"
/codon_start=1
/transl_table=11
/product="PHA synthase"
/protein_id="CAA47035.1"
/db_xref="GI:581529"
/db_xref="SPTREMBL:Q53049"
/translation="MLDBVHKKSLTLDPIGWGPVTSVAGVRNPQAVTAATAEYA
GRKAPAAATRVFNANDPDAMPDPRDRFSDTAOWENPAYFSLQSYLATRAYVE
ELTEAGSDPLQDCAQFANLDFDALPSNFWLWPGVLTAFETGASLLRGARYAA
HDIINRGSLPLKVDSDAFTVCENLAATPGKVVNRNDLIELIQAPQTEQHVAPILAA
PPWLNKYYIIDLAPGRSLAEAWOHGRTVFMISTYRNDESMRHITMDDYVDGJATAL
DVBEITGSPKIEVLISICLGGMAAARAFAGVGDKRVSFTMLNTLLDYISQVGEIG
LLTDPATLDLVEFRMRQGFISGKEMAGSDFMIRAKDLVFNWVSRWMMKGEKPAFDI
LANNEDSTMPAEMHSHYLSRLYGRNELAEGLYVLDQPLNLDIADCTVVGCAINDH
IVPWTSYQAVNLGGDVRYVLTNGGHVAGAVNPCKRKRWFKAVGAPDAESGTPPAD
POWDEAATRYEHSWEDWTANSNKRAGELVAPPMGSTAHPLEDAPGTIVFS"
2547. .2963
/ note="ORF3"
/codon_start=1
/transl_table=11
/protein_id="CAA47036.1"
/db_xref="GI:46401"
/db_xref="SPTREMBL:Q53050"
/translation="MTTATPVDAAVAKTTADAAKANEAAKATADATAAQAIAESA
TAAANQATAEFTKALDSDAARLSSEFNDKVYDAAKQSGNLTVDTYERAVSSLLDLQKEL
ASASPVGWVEDLTAKQVSFARELTATVTTTARDLLK"
2969. .3009
terminator
3041. .>3875
/ note="ORF4"
/codon_start=1
/transl_table=11
/protein_id="CAA47037.1"
/db_xref="GI:46402"
/db_xref="SPTREMBL:Q53051"
/translation="MDALQTVATAAAPFRLGAAAIAGWSSYLATLARRGAHPLDVATD

```



```

/product="Pkbj"
/protein_id="AAF86389.1"
/db_xref="GI:9280391"
/translation="MTAPVKDDPVAHALVAFKTKTRSDPVDRLFAEGGLTSLFAM
ELVYLKTFDVTIAGDPLQLANERTVESMVALVHRLRAVDA"
/gene="fkbk"
/complement(11245..12093)
/gene="fkbk"
/complement(11245..12093)
/gene="fkbk"
/acyl CoA dehydrogenase"
/codon_start=1
/translation="Pkbk"
/protein_id="AAF86390.1"
/db_xref="GI:9280392"

Query Match      16.3%; Score 44.6; DB 1; Length 77534;
Best Local Similarity 53.1%; Pred. No. 17;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 96 ttcatcgcaccgagaaagcccatgagccgagccgctctccacccctctccctccctcc 155
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 46349 TACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 155
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 156 gtcatcctcaccgaagacgattgaagaaatccgcctcacaagccgctcgagtcagt 215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 46289 GCCGAACTCGCGCTTCTGACGACTCCGCGCGCGCGCGCGCGCGCGCGCGCG 215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 216 gagtcggcgatgctcctgctgagcagcgctccaccgccaagtcgctgacgc 274
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 46229 CACCTGACACGAGTGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 274
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AC007789 LOCUS
AC007789 182756 bp DNA PLN 03-DEC-1999
DEFINITION Oryza sativa BAC OSJNBa0049B20 genomic sequence, complete sequence.
AC007789
VERSION AC007789.1 GI:5042437
KEYWORDS HTG.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
Buell, R., Benito, M.-I., Lin, X., Mason, T.M., Umayam, L., Shea, T.P.,
Fujii, C.Y., Shen, M. and Fraser, C.M.
Oryza sativa BAC OSJNBa0049B20 genomic sequence
Unpublished
2 (bases 1 to 182756)
Benito, M.-I.
Direct Submission
Submitted (11-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
3 (bases 1 to 182756)
Benito, M.-I.
Direct Submission
Submitted (15-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 182756)
Benito, M.-I.
Direct Submission
Submitted (18-JUN-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
5 (bases 1 to 182756)
Benito, M.-I.
Direct Submission
Submitted (03-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
Address all correspondence to:
Robin Buell or Maria-Ines Benito
The Institute for Genomic Research
9712 Medical Center Dr.

```

Rockville, MD 20850, USA
e-mail:rbuell@tigr.org or mbenito@tigr.org
BAC clone OSJNBa0049B20 is from *Oryza sativa*.
The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis and Rice EST databases at TIGR, and the maize EST database at Genbank (<http://www.tigr.org/tdb/at/at.html>).

<http://www.tigr.org/tigr/home/tdb/ogi/index.html>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smith, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

FEATURES

source

```

1..182756
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/clone="OSJNBa0049B20"
/complement(join(<179..204,560..775,896..1020,1264..1303,
1587..2987,3531..3610,3722..3766))
/gene="OSJNBa0049B20.1"
/complement(<179..>3766)
/gene="OSJNBa0049B20.1"
/feature="predicted by genscan and genefinder"
/complement(join(179..204,560..775,896..1020,1264..1303,
1587..2987,3531..3610,3722..3766))
/gene="OSJNBa0049B20.1"
/feature="hypothetical protein"
/codon_start=1
/protein_id="AAD38275.1"
/db_xref="GI:5042438"
/translation="MVSSSAATCGPRITAKIDTSVRFGESSINSFAIYICADGK
LIDGSSPPVHTPIATDHSNAAAAAASVRSRRKMEHSPHLEFGAIPKRRDE
RMAAPREATWVILACVPSVSSDGEADGDLAFDWRDPGVSLLTLRQSDSVFSP
APRDFCPDRDDHPYVAVDSAGLLLRGARRSAHDFGPGVALGDFAPSCANDGGYIL
CHATLRMAYLYPCSDVEYRLIYAGNVGIRRTAADGHPRLIAELQIESGNGTHRAT
LLRYSQLGLGGWASTKVNYPGRRSCWGDGVIHAGLWLVDSGLLNCDAFAKP
DMREVPLEGGCKLPYSSDADHAKHRCVNDGELAFVQIHDTAAGRAPSTIMISM
WTLQSDAGEESVSLRHRVDEIDHVTYRTKTMPPRRVPVLAHLHPKELGVVFFQ
ITSNSMFAVDLTVRLVLECKYKMPQLEPMTHSSRWELPHSICRGEDETDG
TSYNVMLYFDHVSNTDELDFNSRDKADELLSTTGLRLEINPRFQELRNATAPKYL
SFVIVKATDAYEASCLMRDFVSHVSDMGDSNIVRVSPSEYEDD"
/complement(1959..1980)
/rpt_family="GC-rich"
2892..2946
/rpt_family="CCG)n"
join(<4106..4327,4875..4935,5026..5411)
/gene="OSJNBa0049B20.2"
<4106..>5411
/gene="OSJNBa0049B20.2"
join(4106..4327,4875..4935,5026..5411)
/gene="OSJNBa0049B20.2"
/codon_start=1
/product="putative DNA-binding protein"
/protein_id="AAD38276.1"
/db_xref="GI:5042439"
/translation="MAACFRCAPSVTAASVPAGPSHLPPDGTALSAFVHDVATASTAR

```

repeat_region

repeat_region

mRNA

gene

CDS

Sat Nov 4 18:11:25 2000

```
ISSAMLOQTWNTSRGTLSPSLSSASPIITTCQCLGWNAGLFLSSSSSSSGRGTAMM
ACLDVEMPEFLRGIDVNWAPAGETTVRGGRGAGAAARRTNSPARPPTLRHLLOPA
TTRTRPAVAADLVYRSGRGGSSSSSRREKRERREKEMKIRYAADMWAPRALF
NFFF"
```

repeat_region

5266..5308

/rpt_family="(CAG)n"

5311..5360

/rpt_family="(GAAA)n"

complement(5841..5891)

/rpt_family="(GA)n"

join(<6090..6238,6898..6944,7302..>7675)

/gene="OSJNBa0049B20.3"

<6090..>7675

/gene="OSJNBa0049B20.3"

/note="predicted by gensean"

join(6090..6238,6898..6944,7302..7675)

/gene="OSJNBa0049B20.3"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD38277.1"

/db_xref="GI:5042440"

```
translation="MHPSTPPALRRRRNPHPAALAEALSSTLKAABEGTVEVGEE
DDDEATYRLILNLYRRKSAKELQSSYHRYLEAGMESWEEYTVODREVRALIDE
IVELYRSPRPDVRERRRRGHRRIVVDDVRRRRRSCMPSSSCVFGVGAFLIV
LMGWIFGHVFLVFFSWILSFVCFNSRH"
```

repeat_region

6582..6665

/rpt_family="AT-rich"

complement(7469..7548)

/rpt_family="(CGA)n"

9207..9483

/note="MITE element, Ditto-like"

complement(9758..9826)

/note="exon predicted by xgrail, quality

excellent_shadowexon"

join(<10005..10220,10323..>10382)

/gene="OSJNBa0049B20.4"

<10005..>10382

/gene="OSJNBa0049B20.4"

/note="predicted by gensean and genefinder"

join(10005..10220,10323..10382)

/gene="OSJNBa0049B20.4"

/note="hypothetical protein"

/codon_start=1

/protein_id="AAD38278.1"

/db_xref="GI:5042441"

```
translation="MSGYGRDVRPLVDKGRGVHTERDRGVHTERDRGGGAASIVE
IIVGPTRRSSMAKATAGRRRRRTYVVDTVLVRKFEKDCIFAPFD"
```

misc_feature

10425..10509

/note="exon predicted by xgrail, quality

marginal"

10664..10857

/note="exon predicted by xgrail, quality

excellent_shadowexon"

complement(10672..10787)

/note="exon predicted by xgrail, quality good_shadowexon"

complement(10932..10989)

/note="exon predicted by xgrail, quality

excellent_shadowexon"

complement(join(<11055..11247,12072..12127,12212..12290,

12568..12707,12811..13096,13176..13318,13421..13621,

13739..13867,14576..14851,14936..15211,15477..16522,

16565..>16724))

/gene="OSJNBa0049B20.5"

complement(<11055..>16724)

/gene="OSJNBa0049B20.5"

/note="contains a ZF potential zinc finger"

complement(join(11055..11247,12072..12127,12212..12290,

12568..12707,12811..13096,13176..13318,13421..13621,

13739..13867,14576..14851,14936..15211,15477..16522,

16565..16724))

/gene="OSJNBa0049B20.5"

/codon_start=1

/product="putative CREB-binding protein"

/protein_id="AAD38279.1"

```
/db_xref="GI:5042442"
/translation="MGAANFAHSLADKPKOMPERLANTIFTSCATLPKCSIDVLH
IGHIKHFSAEVWHHWHYCOILPGDAYONDSOSPSTSGSSSLSAVWDQTTCCSAMR
TLPMDSFTVNGQNLSTNNKSLPTTQGPPLQOYTECBKQOETWSRSLSEOSDQNT
TGRNDLYHAQIHPIYINGEHRKDRICIQMKEKLGHTSDHEGFSREKSNLSNPHVHQOG
FMTNYGACSPVSKTVDRAEQTSNSTVSPASDSSGSKHYPAKRLKVDVPHLVHVN
EMEASKBOQPAANETVYASAEQTSNSTVSPASDSSGSKHYPAKRLKVDVPHLVHVN
GSAVOTEERFRRENSDIEMDKAVQDVLDDQTLSDGSLRARKRRGASVLYALTSDELKH
LCILNHDTSQSKVPTSELLSVEGLPQONTCLNCGMERLLPEPPRFCAKCFKINSTG
SYIVVEENGNDKSSICGRCHLSSAKAKYOKRPSYAEADAEAEWWQCDCKKAWOQI
CALFNPKIVDPEAYTCAKFLKEKDNEDVDSLEPSTILGARELPRLRLSDHLEORLS
ERLVQEROQRIATASGKSVDEPVGVEGLTVRVSSADRTLQVRFKDFKKEIPGEF
PYKSKAILLFQKNEGVDVCLFAMYVOEYGSACFSPNORHVLYLAYIDSVKYFKEI
SREALRTFVYHEILIGVDFCKRGFVSCSIWTCPSYKRDYVLYCHPTIKRMPKSDK
LRSHYQNLVKKVKEGVVVERNTLYDFLOPTNECKTNI SAALPVCNDNDFWPGAEAR
LLEKKDDTSQKKETQLGRLLRVAKRDRKGNLEDILLVHKLGERLRTMKEDFLMLCL
```

Query Match

Best Local Similarity 16.3%; Score 44.6; DB 7; Length 182756;

Matches 92; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 101 cgcacccgagaaagcgcgcacatggaacgcgcgcctccacccctccctccctccctccat 160

Db 156994 CGCCGCCGACCGCGCGCATCGGTGTGGGTGCTTCTGCACTACAGCGCCCGCGCGC 157053

QY 161 cctcaccccaagacgatttgaagaaatcgccgcctacacagccgcgtacgtgagtc 220

Db 157054 GGGCGCGCGCGCGGTGGACGACGCGGGGTCTCGACGAGCGCGGAGGTGGTGT 157113

QY 221 cggcgtggtctcgcgcctaggaacccgcgcctccacccgcgaatgcgcgcga 271

Db 157114 CGCGCGCAGGCTGGCGCGCTGCGCGCGCGACGCGCGCATGTCGAGGA 157164

Search completed: November 4, 2000, 13:37:55

Job time: 17324 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:16 ; Search time 320.8 Seconds
(without alignments)
320.859 Million cell updates/sec

Title: US-09-300-482-311
Perfect score: 274
Sequence: 1 ctacattcttctccacc.....gccaaagtcgcgtcgaccg 274

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_36:*

- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	24.8	1118	18 T87859	DNA encoding ribos
2	44.6	16.3	77536	21 A14651	Nucleotide sequenc
3	43	15.7	701	18 T94500	GAL4 DNA binding d
4	43	15.7	38734	20 Z32020	Human METH1 relate
5	42.6	15.5	1371	21 Z87214	DNA encoding synth
6	42	15.3	771	20 Z12317	Neisseria meningit
7	41.8	15.3	833	15 Q64203	snab gene encoding
8	41.8	15.3	882	19 V64512	M. tuberculosis im
9	41.8	15.3	882	19 V44403	Mycobacterium tube
10	41.8	15.3	882	20 Z19313	M. tuberculosis an
11	41.8	15.3	882	20 Z19101	M. tuberculosis re
12	41.8	15.3	5392	15 Q64201	Sequence compris

c 13	41.6	15.2	290	18	T91500	Mycobacterium tube
c 14	41.6	15.2	290	18	T91436	M. tuberculosis im
c 15	41.6	15.2	290	19	V64483	Mycobacterium tube
c 16	41.6	15.2	290	19	V44375	M. tuberculosis an
c 17	41.6	15.2	290	20	Z19285	M. tuberculosis re
c 18	41.6	15.2	290	20	Z19073	Sorghum bicolor (L
c 19	41.6	15.2	1929	19	V57472	T. thermophilus ga
c 20	41	15.0	3632	17	T06978	Thermus thermophil
c 21	41	15.0	3632	20	X21373	M. tuberculosis im
c 22	40.8	14.9	985	19	V64548	Mycobacterium tube
c 23	40.8	14.9	985	19	V44439	M. tuberculosis an
c 24	40.8	14.9	985	20	Z19349	M. tuberculosis re
c 25	40.8	14.9	985	20	Z19137	Maize T2 cDNA nuc
c 26	40.4	14.7	1236	16	T45060	S. venezuelae deso
c 27	40.4	14.7	1248	21	T87286	S. venezuelae deso
c 28	40.4	14.7	5970	21	Z56003	Contig 002 from co
c 29	40.4	14.7	1244	21	Z87284	S. venezuelae deso
c 30	40.4	14.7	13613	21	Z87319	S. venezuelae deso
c 31	40.2	14.7	13440	19	V52261	Streptococcus pneu
c 32	39.8	14.5	771	20	Z12316	Neisseria meningit
c 33	39.8	14.5	771	20	Z12318	Neisseria gonorrhoe
c 34	39.8	14.5	1574	21	Z48807	Rice inositol 1,3,
c 35	39.8	14.5	1574	21	Z48807	Sequence encoding
c 36	39.4	14.4	3331	21	Z99245	Nucleotide sequenc
c 37	39	14.2	2064	14	Q52638	Streptomyces fradi
c 38	39	14.2	2514	17	T30308	Mouse Sox-9 cDNA.
c 39	38.4	14.0	4466	21	A14663	Nucleotide sequenc
c 40	38.4	14.0	4478	21	A14661	Nucleotide sequenc
c 41	38.4	14.0	4547	21	A14664	Nucleotide sequenc
c 42	38.4	14.0	4571	21	A14662	Nucleotide sequenc
c 43	38.4	14.0	77536	21	A14651	Nucleotide sequenc
c 44	38.2	13.9	396	21	Z34971	Wheat beta-caroten
c 45	38.2	13.9	2291	9	N80309	Entire amylase gen

ALIGNMENTS

RESULT 1

T87859

ID T87859 standard; cDNA; 1118 BP.

XX

AC T87859;

XX

DT 21-APR-1998 (first entry)

XX

DE DNA encoding ribose-5-phosphate isomerase.

XX

KW Ribose-5-phosphate isomerase; screen; inhibitor; herbicidal agent;

KW ss.

XX

OS Spinacia oleracea.

XX

FH Key Location/Qualifiers

FT CDS 25..894

FT /*tag= a

FT /product= ribose-5-phosphate_isomerase

XX

PN WO9737028-A2.

XX

PD 09-OCT-1997.

XX

PF 26-MAR-1997; 97WO-EP01539.

XX

PR 29-MAR-1996; 96DE-4012772.

XX

PA (BADI) BASF AG.

XX

PI Kellermann J, Lerchl J, Schmidt R, Schnarrenberger C;

PI William M;

XX

WPI: 1997-503113/46.

DR P-PSDB; W27505.


```

XX FH Key Location/Qualifiers
XX FT CDS 3..806
XX FT /*tag= a
XX PN WO9816645-A2.
XX PD 23-APR-1998.
XX PF 07-OCT-1997; 97WO-US18214.
XX PR 13-MAR-1997; 97US-0818111.
XX PR 11-OCT-1996; 96US-0729622.
XX PF (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
XX PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1998-251292/22.
XX DR P-PSDB; W64359.
XX PS New isolated Mycobacterium tuberculosis polypeptides and DNA - used
XX PS to develop products for the detection of M. tuberculosis infection
XX PS and diagnosis of tuberculosis
XX PS Claim 4; Page 153-154; 250pp; English.
XX CC This DNA sequence codes for an antigenic portion (see W64359) of
XX CC Mycobacterium tuberculosis antigen TBH-29. It was isolated from
XX CC a M. tuberculosis strain H37Kv genomic library. The invention
XX CC relates to compositions and methods for diagnosing tuberculosis.
XX CC It provides polypeptides (see W64291-W64379) comprising an
XX CC antigenic portion of a soluble M. tuberculosis antigen, or an
XX CC immunogenic portion of an M. tuberculosis antigen, as well as DNA
XX CC sequences encoding such polypeptides, recombinant expression
XX CC vectors and transformed or transfected host cells. Also claimed
XX CC are methods and diagnostic kits for detecting M. tuberculosis
XX CC infection in a patient using these polypeptides, antibodies or
XX CC oligonucleotide probes and primers, for the diagnosis of
XX CC tuberculosis.
XX SQ Sequence 882 BP; 141 A; 355 C; 277 G; 109 T; 0 other;

Query Match 15.3%; Score 41.8; DB 19; Length 882;
Best Local Similarity 51.9%; Pred. No. 0.058;
Matches 94; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 92 ccatttcacgccaccgagaaagccgcatggagccggcctctccaccctctctccc 151
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 388 ccattccaccgacgcgcgacgacgcgggtgaccacgtcgagcagcgcgcgacca 447
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 ctccgtcatctccaccgagacgatttgaagaaatccgcgcctacaaagccgtcgagta 211
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 448 cgccgcgaccacgcgcggtgaccacgcgcgaagacgcgcgcgcgcgcgcgcgcgc 507
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 cgtgagtcgcgcgtgctctggcctagggcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 271
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 508 cgccgcgaacgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 271
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 272 c 272
   |
Db 568 c 568

RESULT 10
ID Z19313
XX Z19313 standard; DNA; 882 BP.
AC Z19313;
XX 05-NOV-1999 (first entry)
DT

```

```

XX DE M. tuberculosis antigen TBH-29 DNA sequence.
XX KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
XX KW immunotherapy; diagnosis; immunisation; vaccine; infection;
XX KW immune response; skin test; ss.
XX OS Mycobacterium tuberculosis.
XX PN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US03268.
XX PR 05-MAY-1998; 98US-0072967.
XX PR 18-FEB-1998; 98US-0025197.
XX PF (CORI-) CORIXA CORP.
XX PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
XX PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX DR WPI; 1999-527409/44.
XX DR P-PSDB; Y39156.
XX PS New antigens from Mycobacterium tuberculosis useful in diagnostic
XX PS skin tests and protective or therapeutic vaccines or compositions
XX PS Claim 4; Page 141; 299pp; English.
XX CC The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
XX CC are vaccines and fusion protein containing M. tuberculosis Ag's.
XX CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
XX CC other polypeptides fragments, can be used in pharmaceutical compositions
XX CC or vaccines to generate a protective or therapeutic immune response to
XX CC M. tuberculosis and as reagents in skin tests for diagnosis of
XX CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
XX CC by, T, B or natural killer cells and/or macrophages in
XX CC tuberculosis-immune subjects. Z19249 to Z19460 and Y39083 to Y39225
XX CC are used in the exemplification of the present invention.
XX SQ Sequence 882 BP; 141 A; 355 C; 277 G; 109 T; 0 other;

Query Match 15.3%; Score 41.8; DB 20; Length 882;
Best Local Similarity 51.9%; Pred. No. 0.058;
Matches 94; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 92 ccatttcacgccaccgagaaagccgcatggagccggcctctccaccctctctccc 151
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 388 ccattccaccgacgcgcgacgacgcgggtgaccacgtcgagcagcgcgcgcgcgcgcgcgcgc 447
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 ctccgtcatctccaccgagacgatttgaagaaatccgcgcctacaaagccgtcgagta 211
   |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 448 cgccgcgaccacgcgcggtgaccacgcgcgaagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 507
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 cgtgagtcgcgcgtgctctggcctagggcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 271
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 508 cgccgcgaacgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 271
   || | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 272 c 272
   |
Db 568 c 568

RESULT 11
ID Z19101
XX Z19101 standard; DNA; 882 BP.
AC Z19101;
XX

```

108

100

100

PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14674.
XX
PR 12-JUL-1996; 96US-0680574.
PR 01-SEP-1995; 95US-0523436.
PR 22-SEP-1995; 95US-0533634.
PR 22-MAR-1996; 96US-0620874.
PR 05-JUN-1996; 96US-0659683.
XX
PA (CORI-) CORIXA CORP.
XX

PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
PI Twardzik DR, Vedvick TH;
PI WPI; 1997-192903/17.
XX

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT useful in vaccines for prevention or treatment of tuberculosis, also
PT for diagnosis
XX

PS Claim 4; Page 78; 168pp; English.
XX

CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC its variant differing only in conservative substitutions and/or
CC modifications). The present sequence encodes a specifically claimed
CC M.tuberculosis antigen, TBm-7. The immunogenic protein, and fusion
CC proteins containing one or more of the proteins or one of the proteins
CC plus ESAT-6, are useful in vaccines, preferably when formulated with a
CC non-specific adjuvant, to induce an immune response against
CC M.tuberculosis (for treatment or prevention).
XX

SQ Sequence 290 BP; 30 A; 83 C; 142 G; 33 T; 2 other;

Query Match 15.2%; Score 41.6; DB 18; Length 290;
Best Local Similarity 51.6%; Pred. No. 0.052;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 89 cccccatttcacccacgagaaagccgacgagcgcctcctccacccctcctc 148
Db 228 CGCCTCGCCACCCCTTACCCCGCGTTCGCCACGAGCGCGCGTTCGCCCGCG 169
QY 149 cccctccgtcatctcaccacgaacgatttgaagaaatcgcgcctacaaggccgtga 208
Db 168 CACCGCCGCGCGCGCAAGACCGCGTTCGCCCGCGTTCGCCCGCGCGCGCGCT 109
QY 209 gtacgtgagtcgagtcgctcctcctcctcctcctcctcctcctcctcctcctc 268
Db 108 GCCCAGCGAGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49
QY 269 cgac 272
Db 48 CGAC 45

RESULT 14
T91436/C
ID T91436 standard; DNA; 290 BP.
XX
AC T91436;
XX
DT 13-JAN-1998 (first entry)
XX

XX Mycobacterium tuberculosis antigen TBm-7 encoding DNA.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis; ss.
XX

OS Mycobacterium tuberculosis.
XX

PN WO9709429-A2.

XX 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US14675.
XX
PR 12-JUL-1996; 96US-0680573.
PR 01-SEP-1995; 95US-0523435.
PR 22-SEP-1995; 95US-0532136.
PR 22-MAR-1996; 96US-0620280.
PR 05-JUN-1996; 96US-0658800.
XX
PA (CORI-) CORIXA CORP.
XX

PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
PI Twardzik DR, Vedvick TH;
PI WPI; 1997-192904/17.
XX

PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
PT - useful for diagnosis of M. tuberculosis infection
PT
XX

PS Claim 4; Page 82-83; 190pp; English.
XX

CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC its variant differing only in conservative substitutions and/or
CC modifications). The present sequence encodes a specifically claimed
CC M.tuberculosis antigen, TBm-7. The immunogenic polypeptide can be
CC used to diagnose M.tuberculosis infection by forming complexes with
CC specific antibodies in the sample. Fragments of DNA encoding the
CC immunogenic polypeptide can be used as diagnostic primers or probes
CC and agents that bind to the antigen, especially monoclonal antibodies
CC or equivalent polyclonal antibodies, are also used for diagnosis.
XX

SQ Sequence 290 BP; 30 A; 83 C; 142 G; 33 T; 2 other;

Query Match 15.2%; Score 41.6; DB 18; Length 290;
Best Local Similarity 51.6%; Pred. No. 0.052;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 89 cccccatttcacccacgagaaagccgacgagcgcctcctccacccctcctc 148
Db 228 CGCCTCGCCACCCCTTACCCCGCGTTCGCCACGAGCGCGCGTTCGCCCGCG 169
QY 149 cccctccgtcatctcaccacgaacgatttgaagaaatcgcgcctacaaggccgtga 208
Db 168 CACCGCCGCGCGCGCAAGACCGCGTTCGCCCGCGTTCGCCCGCGCGCGCGCT 109
QY 209 gtacgtgagtcgagtcgctcctcctcctcctcctcctcctcctcctcctcctc 268
Db 108 GCCCAGCGAGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49
QY 269 cgac 272
Db 48 CGAC 45

RESULT 15
V64483/C
ID V64483 standard; DNA; 290 BP.
XX
AC V64483;
XX
DT 27-JAN-1999 (first entry)
XX

XX M. tuberculosis immunogenic polypeptide TBm-7 DNA.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis; ss.
XX

OS Mycobacterium tuberculosis.
XX

RESULT 7

; GENERAL INFORMATION:
 ; APPLICANT: Nakamura, Etsuo
 ; APPLICANT: Tsuzuki, Hiroshige
 ; APPLICANT: Kitadokoro, Kengo
 ; APPLICANT: Shin, Nasaru
 ; APPLICANT: Teraoka, Hiroshi
 ; TITLE OF INVENTION: No. 5665586el Protease
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: James F. Haley, Jr., Fish & Neave
 ; STREET: 1251 Avenue Of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10020-1104

```

1  COMPUTER READABLE FORM:
2
3  MEDIUM TYPE: Diskette, 3.50 inch, 720kb storage
4
5  COMPUTER: IBM Compatible
6
7  OPERATING SYSTEM: MS Dos 5.0
8
9  SOFTWARE: WordPerfect 5.0
10
11 CURRENT APPLICATION DATA:
12
13 APPLICATION NUMBER: US/08/343,428
14
15 FILING DATE: 18-NOV-1994
16
17 CLASSIFICATION: 435
18
19 PRIORITY APPLICATION DATA:
20
21 APPLICATION NUMBER: PCT/J93/00592
22
23 FILING DATE: 30-APR-1993
24
25 APPLICATION NUMBER: JAPAN 4-126511
26
27 FILING DATE: 19-MAY-1992
28
29 ATTORNEY/AGENT INFORMATION:
30
31 NAME: Haley, Jr., James F.
32
33 REGISTRATION NUMBER: 27794
34
35 REFERENCE/DOCKET NUMBER: SHGN-7
36
37 TELECOMMUNICATION INFORMATION:
38
39 TELEPHONE: (212)596-9000
40
41 TELEFAX: (212)596-9090
42
43 INFORMATION FOR SEQ ID NO: 1:
44
45 SEQUENCE CHARACTERISTICS:
46
47 LENGTH: 2064
48
49 TYPE: nucleic acid
50
51 STRANDEDNESS: double
52
53 TOPOLOGY: linear
54
55 MOLECULE TYPE: Genomic DNA
56
57 ORIGINAL SOURCE:
58
59 ORGANISM: Streptomyces fradiae
60
61 FEATURE:

```

```

; LOCATION: 378..383
; IDENTIFICATION METHOD: by experiment
; NAME/KEY: CDS
; LOCATION: 435..1505
; IDENTIFICATION METHOD: by experiment
; NAME/KEY: sig peptide
; LOCATION: 435..944
; IDENTIFICATION METHOD: by experiment
; --08-343-428-1
US-08-343-428-1

Query Match 14.2% Score 39; DB 1; Length 2064;
Best Local Similarity 55.6% Pred. No. 0.13;
Matches 75; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

Oy 197 caagcgctgcgtagtcgtggagtccggcgaaggctcctcggcctaggacacggctcacgcg 256
 ||| |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Dδ 617 CAGCGCCGGCACCTTACGTGGAACGCCACCGGCACAATCTGTCGTCACGTCACCACCGA 676

```

QY 257 caagcatgccgtcga 271
Db 677 GGCCGGCGCGGCCAA 691

RESULT 10
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTgpt-F1s
US-08-232-463-14

Query Match 14.2%; Score 39; DB 1; Length 7218;
Best Local Similarity 3.2%; Pred. No. 0.17;
Matches 6; Conservative 118; Mismatches 63; Indels 0; Gaps 0;

QY 1 ctatattcttctccaccacacatttgaacctctgggactggctagcttcaa 60
Db 1257 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1316
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 61 cetttaacattaacatggccattccctacccttctcatgccaccgagaaagcgcca 120
Db 1317 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1376
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 121 tggacggcggtctctccaccctctccctccgtctatctcaccacacagattga 180
Db 1377 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1436
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 181 agaaat 187
Db 181 | | | | |

US-08-042-747A-7
; Sequence 7, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042,747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2943 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 87..2744
US-08-042-747A-7

Query Match 13.7%; Score 37.6; DB 1; Length 2943;
Best Local Similarity 49.5%; Pred. No. 0.31;
Matches 97; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 79 ccatttcctacccccatttctatcgccaccgagaaagccgcatggagcgccctctcc 138
Db 1102 CCAAGCGCCCTCGGTGTGCACGCTGACCAAGTGGCGGGAGGTGGACGAGATGCTCGCGG 1161
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 139 accctctccctccgtctatctcaccacacagcatttgaagaaaaatcgccgctaca 198
Db 1162 CCGAGTAGGCGCCCTCGTTCGGCTTCTCTCGCGCCCTCTCGACACACCTTCACCGCCA 1221
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 199 aggcctcgagtagtgagtcggcgatggtctctggtcctagcaccggtccacgcca 258
Db 1222 ACCGACGAGTAGTACGCTCTCGCGGCTCGACCTCGCGGACTCGTCGGGCGGAGGCC 1281
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 259 agcatgcgtgcaccg 274
Db 1282 GCGAGGCGGTGGACCG 1297
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:13 ; Search time 4352.3 Seconds
(without alignments)
389.240 Million cell updates/sec

Title: US-09-300-482-311

Perfect score: 274

Sequence: 1 ctacattcttctccacc.....gccaaagatgcgtgacccg 274

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estrol:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estrol10:*
85: em_estrol11:*
86: em_estrol12:*
87: em_estrol13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

TITLE -ESTs from roots of Medicago truncatula after inoculation with
Phytophthora medicaginis
JOURNAL Unpublished (1999)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612-625-7219
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name: M251150e ; TIGR sequence name: MTHAP91TK ; More
information, including clone ordering, is available at. .
'http://chrysis.tamu.edu/medicago/
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers

FEATURES source

1..563
/organism="Medicago truncatula"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="pDSIR-24P14"
/clone_lib="DSIR"
/tissue_type="roots infected with Phytophthora medicaginis"
/dev_stage="roots harvested at 10 days post inoculation with Phytophthora medicaginis"
/lab_host="E. coli strain SOLR"
/note="vector: pBluescript SK +/-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells. Note: EST may be of fungal origin."
BASE COUNT 140 a 193 c 72 g 158 t
ORIGIN

Query Match 48.8%; Score 133.6; DB 23; Length 563;
Best Local Similarity 80.8%; Pred. No. 1.3e-25;
Matches 173; Conservative 0; Mismatches 29; Indels 12; Gaps 1;
QY 73 acatggccattccctacccatttccatgcacccgagaaagccgcatgagccgccc 132
Db 226 AGATGGCTATTCCTTGCCTCCATTTTCATCTCTCCGCAAAAGCCGCATGGAAGCCGTC 285
QY 133 tctccaccctctctccccc-----ctcgcgtatctctcaccacgacgattga 180
Db 286 TCCTCATCCCTTCCTCCCATCTTCTCCGTCCTCGTCATCTATCCCAAGACGATTTGA 345
QY 181 agaaatcgccgtacagccgctgagtagtgagtcgagtcgcatggtctcgcgctag 240
Db 346 AGAAATCGCGCTTACAAAGCCGTCGAATACGTCGAATCGGTATGTTCTAGGCTCG 405
QY 241 gcacggctccacgcgaagcatccgctgaccg 274
Db 406 GAACCGGTTCAACGCCCAACACGCGGTCGCCG 439

RESULT 3
AW559587 668 bp mRNA EST 10-MAY-2000
LOCUS EST314635 DSIR Medicago truncatula cDNA clone pDSIR-24115, mRNA
DEFINITION sequence.
ACCESSION AW559587
VERSION AW559587.1 GI:7205013
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

REFERENCE AUTHORS TITLE JOURNAL COMMENT

Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 668)
Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, G.S., Peng
H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,
Holt, I.E., and Town, C.M.
ESTs from roots of Medicago truncatula after inoculation with
Phytophthora medicaginis
Unpublished (1999)
Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612-625-7219
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name: M250873e ; TIGR sequence name: MTHAM56TK ; More
information, including clone ordering, is available at. .
'http://chrysis.tamu.edu/medicago/
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers

FEATURES source

1..668
/organism="Medicago truncatula"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="pDSIR-24115"
/clone_lib="DSIR"
/tissue_type="roots infected with Phytophthora medicaginis"
/dev_stage="roots harvested at 10 days post inoculation with Phytophthora medicaginis"
/lab_host="E. coli strain SOLR"
/note="vector: pBluescript SK +/-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells. Note: EST may be of fungal origin."
BASE COUNT 162 a 202 c 108 g 196 t
ORIGIN

Query Match 48.8%; Score 133.6; DB 23; Length 668;
Best Local Similarity 80.8%; Pred. No. 1.4e-25;
Matches 173; Conservative 0; Mismatches 29; Indels 12; Gaps 1;
QY 73 acatggccattccctacccatttccatgcacccgagaaagccgcatgagccgccc 132
Db 208 AGATGGCTATTCCTTGCCTCCATTTTCATCTCTCCGCAAAAGCCGCATGGAAGCCGTC 267
QY 133 tctccaccctctctccccc-----ctcgcgtatctctcaccacgacgattga 180
Db 268 TCCTCATCCCTTCCTCCCATCTTCTCCGTCCTCGTCATCTATCCCAAGACGATTTGA 327
QY 181 agaaatcgccgtacagccgctgagtagtgagtcgagtcgcatggtctcgcgctag 240
Db 328 AGAAATCGCGCTTACAAAGCCGTCGAATACGTCGAATCGGTATGTTCTAGGCTCG 387
QY 241 gcacggctccacgcgaagcatccgctgaccg 274
Db 388 GAACCGGTTCAACGCCCAACACGCGGTCGCCG 421

RESULT 4
AI900328 428 bp mRNA EST 06-DEC-1999
LOCUS sc04b07.y1 Gm-cl012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl012-1166 5' similar to SW-RP1A_ECOLI P27252 RIBOSE 5-PHOSPHATE
ISOMERASE A ; mRNA sequence.
ACCESSION AI900328

[illegible]

Query Match	32.1%;	Score 88;	DB 18;	Length 382;
Best Local Similarity	80.5%;	Pred. No. 1.6e-13;		
Matches 103;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;
145	ccctcccccctccgctcatctctca	cccccaagcagattgga	aaaaatcgccgcgtctacaagggccg	204
210	CCATTCGTCGTCATCACCCCTC	ACCCACCAAGACGACGCTC	AAGAGACTCGCGCGCCGACAAAGSCCG	269
205	tcgagtcagtggagtcggcgatgg	tccctcggcctaggcacccgctccacgcgcgaacgatg	264	
270	TCGACTACGTCACAGTCGGCG	ATGTCCTCGGCCTCGGCACCGCGGCTCCACCGCGCCCTTCG	329	
265	ccgtcgac	272		
330	TCGTCGCC	337		

RESULT	14
AVAV07718	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	

AV407718 389 bp mRNA EST 23-MAY-2000
AV407718 Lotus japonicus young plants (two-week old) Lotus
japonicus cDNA clone MWL029c05_r 5', mRNA sequence.
AV407718
AV407718.1 GI:7720572
EST.
Lotus japonicus.
Lotus japonicus
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Lotus.
1 (bases 1 to 389)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
Generation of 7137 non-redundant expressed sequence tags from a
legume, *Lotus japonicus*
DNA Res. 7 (2), 127-130 (2000)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:34:27 ; Search time 4075.18 Seconds
(without alignments)
415.840 Million cell updates/sec

Title: US-09-300-482-298
Perfect score: 388

Sequence: 1 ggagaaagaagaagaagatg.....ttaagcctgggaccccgct 388

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_pl1: *
8: gb_pl2: *
9: gb_pr1: *
10: gb_pr2: *
11: gb_pr3: *
12: em_fun: *
13: em_hum1: *
14: em_hum2: *
15: em_in: *
16: em_om: *
17: em_or: *
18: em_ov: *
19: em_pat: *
20: em_ph: *
21: em_pl: *
22: em_ro: *
23: em_sts: *
24: em_sy: *
25: em_un: *
26: em_vi: *
27: gb_htg1: *
28: gb_htg2: *
29: gb_in1: *
30: gb_in2: *
31: em_ba1: *
32: em_ba2: *
33: em_hum3: *
34: em_hum4: *
35: gb_pr4: *
36: gb_htg3: *
37: gb_htg4: *
38: gb_htg5: *
39: gb_htg6: *
40: gb_htg7: *
41: em_htg1: *
42: em_htg2: *
43: em_htg3: *

44: em_hum5: *
45: gb_pl3: *
46: gb_pr5: *
47: gb_htg8: *
48: gb_htg9: *
49: gb_htg10: *
50: gb_htg11: *
51: gb_htg12: *
52: gb_htg13: *
53: gb_htg14: *
54: gb_in3: *
55: gb_htg15: *
56: gb_htg16: *
57: gb_htg17: *
58: em_htg4: *
59: em_htg5: *
60: em_htg6: *
61: em_htg7: *
62: em_hum6: *
63: gb_htg18: *
64: gb_htg19: *
65: gb_ba3: *
66: em_htg8: *
67: em_htg9: *
68: em_htg10: *
69: em_htg11: *
70: em_htg12: *
71: em_htg13: *
72: em_htg14: *
73: em_htg15: *
74: em_htg16: *
75: em_htg17: *
76: em_htg18: *
77: em_htg19: *
78: em_htg20: *
79: em_htg21: *
80: em_htg22: *
81: em_htg23: *
82: gb_pr6: *
83: gb_pr7: *
84: gb_htg20: *
85: gb_htg21: *
86: gb_htg22: *
87: gb_htg23: *
88: gb_ro: *
89: gb_sts1: *
90: gb_sts2: *
91: gb_sy: *
92: gb_un: *
93: gb_vil: *
94: gb_vil2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
1	239.8	61.8	930	7	AF189365	AF189365	Oryza sat
2	85.2	22.0	14229	53	AC057311	AC057311	Giardia
3	85.2	22.0	14229	53	AE001286	AE001286	Chlamydia
4	80.8	20.8	122925	85	AL360001	AL360001	Homo sapi
5	79.2	20.4	153804	27	AC006280	AC006280	Plasmodiu
6	79.2	20.4	196149	27	AC004709	AC004709	Plasmodiu
7	79	20.4	911	45	SCPOS18	X83571	S.cerevisia
8	79	20.4	1328	45	SCYJL121C	249396	S.cerevisia
9	78.6	20.3	5960	2	RCU23145	U23145	Rhodobacter
10	77.6	20.0	12939	1	AE002306	AE002306	Chlamydia
11	76.4	19.7	10150	1	AE002217	AE002217	Chlamydia
12	76.4	19.7	12973	1	AE001604	AE001604	Chlamydia

QELIOSIKAKGMRPCVSLRPGTPEVEPPLVEAENPVELVMTVEPGFGGKQWPEM
MKVRLALRRKPPSLDIEVDGGLGISTIDVAASAGANCIVAGSSIFGAAPCEVISALR
KSVESGQNKs*

BASE COUNT 230 a 206 c 256 g 238 t
ORIGIN

Query Match 61.8%; Score 239.8; DB 7; Length 930;
Best Local Similarity 78.8%; Pred. No. 6e-55; Indels 0; Gaps 0;
Matches 286; Conservative 0; Mismatches 77;

QY 26 gacacgcaaaatagctccctcgatgctctctccgacttcgcaatttgcttcgagggc 85
DB 49 GCGGGCGAAGATAGCGCGTGCATGCTCTGCTCGGACTTCGCCAACCTCGCGCGGAGGC 108
QY 86 tcagcgcgactcccaacttcgcccgcgattggtccacatgagacatcatgagggcattt 145
DB 109 CGACCGCATGTCGCGCTCGCGCGGACTGCGCTCCATGGACATCATGACGGGCATTT 168
QY 146 tglccccaatttaactattgctccgctccagttattgaaagtttgagaaagcacacaaaggc 205
DB 169 TGTTCCTAATCTTACATATTGAGCTCCAGTGATTCAGAGCTTCAGGAAGCACACCAAGGC 228
QY 206 atattggattgacacttatggtacaaatctctctgattgattgtaaccttggaacaa 265
DB 229 ATATTGGACTGCCATCTTATGTTGACCAATCCTTCGGATTATGTAGAACATTAGCAAA 288
QY 266 agctgggtctctggttttaccatttcacatgtagagacataaaagataactggaaagact 325
DB 289 AGCTGGTGCCTTCAGGTTTCACATATCCATATAGAGTATCCAGAGACAAATTTGGCAAGACT 348
QY 326 tatccaaagaatcaagtcacatgcatgattctctggtgtagcattaaagcctgggacccc 385
DB 349 CATCAAAGTATCAAAAGAGGATATCGACCGGGTGTATCATTTAGAGCGGACGACTCC 408
QY 386 cgt 388
DB 409 TGT 411

RESULT 2
AF189365 930 bp mRNA 10-JAN-2000
LOCUS Oryza sativa D-ribulose-5-phosphate 3-epimerase mRNA, complete cds.
DEFINITION AF189365
ACCESSION AF189365.1 GI:6007802
KEYWORDS
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 930)
AUTHORS Kopriva, S., Koprivova, A., and Suss, K.H.
TITLE Identification, cloning, and properties of cytosolic D-ribulose-5-phosphate 3-epimerase from higher plants
J. Biol. Chem. 275 (2), 1294-1299 (2000)
JOURNAL 20092904
MEDLINE 2 (bases 1 to 930)
REFERENCE Kopriva, S., Koprivova, A., and Suss, K.H.
AUTHORS Kopriva, S., Koprivova, A., and Suss, K.H.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-1999) Institute of Forest Biology and Tree Physiology, Am Flughafen 17, Freiburg im Breisgau 79085, Germany
FEATURES
Source Location/Qualifiers
db_xref="taxon:4530"
EC_number="5.1.3.1"
note="cytosolic isoform"
codon_start=1
product="D-ribulose-5-phosphate 3-epimerase"
protein_id="AA01048.1"
db_xref="GI:6007803"
translation="MAAAAKIAPSMILSSDFANLAAEADRMVRLGADWLHMDIMGH
FVPLNLTICAPVQSLRKHXYLDCHLWNTNPSDYVEFLAKAGASGFTFIEVSRDWN

ALIGNMENTS

AP002545 Chlamydomonas reinhardtii
AF070943 Expressio
AF070942 Expressio
L42328 Spinacia ol
AF070941 Spinacia
AC022355 Arabidops
U97573 Treponema p
AE001263 Treponema
Z98979 S.pombe chr
Z50098 S.tuberosum
D90911 Synecocyst
AC010797 Arabidops
AF047444 Oryza sat
Y13937 Bacillus su
Z99112 Bacillus su
AF269471 Staphyloc
AF270217 Staphyloc
AF270200 Staphyloc
AE000482 Escherich
Continuation (2 of
AC034250 Homo sapi
U00006 E. coli chr
AF015274 Arabidops
AE001811 Thermotog
AC019942 Drosophil
U28991 Caenorhabdi
AC005641 Drosophil
AC005450 Drosophil
AE003840 Drosophil
Z80108 Mycobacteri
AE003874 Xylella f
Z78161 M.musculus
AE004329 Vibrio ch


```
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA332019
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 103697 bases at least Q40
Consensus quality: 110797 bases at least Q30
Consensus quality: 115383 bases at least Q20
Insert size: 120025; sum-of-contigs
Insert size: 170370; 12.7% error; agarose-fp
Quality coverage: 0.00x in Q20 bases; sum-of-contigs Quality
Coverage: 0.00x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3339: contig of 3339 bp in length
* 3340 3439: gap of 100 bp
* 3440 6321: contig of 2882 bp in length
* 6322 6421: gap of 100 bp
* 6422 10506: contig of 4085 bp in length
* 10507 10606: gap of 100 bp
* 10607 16241: contig of 5635 bp in length
* 16242 16341: gap of 100 bp
* 16342 19584: contig of 3243 bp in length
* 19585 19684: gap of 100 bp
* 19685 22008: contig of 2324 bp in length
* 22009 22108: gap of 100 bp
* 22109 24532: contig of 2424 bp in length
* 24533 24632: gap of 100 bp
* 24633 27409: contig of 2777 bp in length
* 27410 27509: gap of 100 bp
* 27510 31499: contig of 3990 bp in length
* 31500 31599: gap of 100 bp
* 31600 36750: contig of 5151 bp in length
* 36751 36850: gap of 100 bp
* 36851 39844: contig of 2994 bp in length
* 39845 39944: gap of 100 bp
* 39945 42665: contig of 2721 bp in length
* 42666 42765: gap of 100 bp
* 42766 45870: contig of 3105 bp in length
* 45871 45970: gap of 100 bp
* 45971 48390: contig of 2420 bp in length
* 48391 48490: gap of 100 bp
* 48491 51340: contig of 2850 bp in length
* 51341 51440: gap of 100 bp
* 51441 54723: contig of 3283 bp in length
* 54724 54823: gap of 100 bp
* 54824 56829: contig of 2006 bp in length
* 56830 56929: gap of 100 bp
* 56930 60492: contig of 3563 bp in length
* 60493 60592: gap of 100 bp
* 60593 67232: contig of 6640 bp in length
* 67233 67332: gap of 100 bp
* 67333 70877: contig of 3545 bp in length
* 70878 70977: gap of 100 bp
* 70978 73415: contig of 2438 bp in length
* 73416 73515: gap of 100 bp
* 73516 80763: contig of 7248 bp in length
* 80764 80863: gap of 100 bp
* 80864 83214: contig of 2351 bp in length
* 83215 83314: gap of 100 bp
* 83315 87723: contig of 4409 bp in length
* 87724 87823: gap of 100 bp
* 87824 93411: contig of 5588 bp in length
* 93412 93511: gap of 100 bp
*
* 93512 97223: contig of 3712 bp in length
* 97224 97323: gap of 100 bp
* 97324 103588: contig of 6265 bp in length
* 103589 103688: gap of 100 bp
* 103689 109134: contig of 5446 bp in length
* 109135 109234: gap of 100 bp
* 109235 112265: contig of 3031 bp in length
* 112266 112365: gap of 100 bp
* 112366 122925: contig of 10560 bp in length.
*
FEATURES
      source
      1. 122925
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="10"
      /clone_lib="RPCI-11.2"
      1. 3339
      /note="assembly_fragment:00165.0"
      3440. 6321
      /note="assembly_fragment:00185.0"
      6422. 10506
      /note="assembly_fragment:00209.0"
      10607. 16241
      /note="assembly_fragment:00218.0"
      16342. 19584
      /note="assembly_fragment:00304.0"
      19685. 22008
      /note="assembly_fragment:00308.0"
      22109. 24532
      /note="assembly_fragment:00374.0"
      24633. 27409
      /note="assembly_fragment:00451.0"
      27510. 31499
      /note="assembly_fragment:00459.0"
      31600. 36750
      /note="assembly_fragment:00480.0"
      36851. 39844
      /note="assembly_fragment:00486.0"
      39945. 42665
      /note="assembly_fragment:00514.0"
      42766. 45870
      /note="assembly_fragment:00595.0"
      45971. 48390
      /note="assembly_fragment:00635.0"
      48491. 51340
      /note="assembly_fragment:00646.0"
      51441. 54723
      /note="assembly_fragment:00658.0"
      54824. 56829
      /note="assembly_fragment:00660.0"
      56930. 60492
      /note="assembly_fragment:00701.0"
      60593. 67232
      /note="assembly_fragment:00789.0"
      67333. 70877
      /note="assembly_fragment:00838.0"
      70978. 73415
      /note="assembly_fragment:00912.0"
      73516. 80763
      /note="assembly_fragment:01027.0"
      80864. 83214
      /note="assembly_fragment:01043.0"
      83315. 87723
      /note="assembly_fragment:01079.0"
      87824. 93411
      /note="assembly_fragment:01273.0"
      93512. 97223
      /note="assembly_fragment:01345.0"
      97324. 103588
      /note="assembly_fragment:01357.0"
      103689. 109134
      /note="assembly_fragment:01384.0"
      109235. 112265
      misc_feature
```



```

* 47832 179129: contig of 131298 bp in length
* 179130 179329: gap of unknown length
* 179330 196149: contig of 16820 bp in length.
FEATURES             Location/Qualifiers
     source           1..196149
                     /organism="Plasmodium falciparum"
                     /db_xref="taxon:5833"
                     /chromosome="12"
BASE COUNT      80057 a 19753 c 18800 g 77138 t 401 others
ORIGIN
20.4%; Score 79.2; DB 27; Length 196149;
Best Local Similarity 55.4%; Pred. No. 5.1e-11;
Matches 175; Conservative 0; Mismatches 138; Indels 3; Gaps 1;

QY 25 tgacacgaaataagctccttcgactgctctcgcacttcgcaattggcttcagg 84
Db 13506 TAAAGCTATAATTGCCCTTCAGTACTTCGATCCCAATATAAGCAATAGCTGAAGAA 13565

QY 85 ctgagcgatgctccacttcgagcgagctgctcgcacatggacatcagtgatggcaatt 144
Db 13566 CACAACGATGGAACTTTGGGTGCGAATGATCCATTAGATGTTATGATATGCAAT 13625

QY 145 ttgtcccaatttaactattggcgtccagttattgaaagtttgagaagcacacaaagg 204
Db 13626 TTGTCCTAATTATCTTTGGTCCACCTGTTATTAAATTTAAAAAATATACAAAA 13685

QY 205 catatttgatgacacattatggttacaaatcctcttgattatgttgaaaccttgcaa 264
Db 13686 GTATTTTGTGATGACACTTAATGGTTGATATCCAGAAATATATGACCATTTGTA 13745

QY 265 aagctggtgctctggttttcatcttcagtagagacataaagataactggaaagaaac 324
Db 13746 AAAC---ATCAATCACTAATCTTTTCATTTTGAGAGCATTAATGAAGATACCGA 13802

QY 325 ttatccaaagaatcaa 340
Db 13803 GTATACAATTAGCCAA 13818

RESULT 7
LOCUS      SCPOS18          911 bp      DNA      PLN      11-MAR-1997
DEFINITION S.cerevisiae RPE gene.
ACCESSION  X83571
VERSION    X83571.1 GI:609673
KEYWORDS  rep gene; Ribulose-5-phosphate-epimerase.
SOURCE    baker's yeast.
ORGANISM  Saccharomyces cerevisiae
           Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
           Saccharomycetaceae; Saccharomycetes.
REFERENCE  1 (bases 1 to 911)
AUTHORS   Juhnke, H.
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 911)
AUTHORS   Juhnke, H.
TITLE     Direct Submission
SUBMITTED (19-DEC-1994) H. Juhnke, Institut fuer Mikrobiologie,
Johann Wolfgang Goethe-Universitaet, Marie-Curie-Strasse 9,
Gebäude N250, AG Entian, 60439 Frankfurt, FRG
3 (bases 1 to 911)
Miosga, T. and Zimmermann, F.K.
Genetic and functional analysis of the Saccharomyces cerevisiae
gene EPI1 encoding ribulose-5-phosphate 3-epimerase
JOURNAL  yeast 11, 500-500 (1995)
REFERENCE 4 (bases 1 to 911)
AUTHORS   Miosga, T. and Zimmermann, F.K.
TITLE     Cloning and characterization of the first two genes of the
non-oxidative part of the Saccharomyces cerevisiae
pentose-phosphate pathway
JOURNAL  Curr. Genet. 30 (5), 404-409 (1996)
MEDLINE  97074231

FEATURES             Location/Qualifiers
     source           1..911
                     /organism="Saccharomyces cerevisiae"
                     /isolate="clone 6"
                     /db_xref="taxon:4932"
                     /clone_lib="plasmid library"
                     120..836
                     /gene="rPE"
                     120..836
                     /gene="rPE"
                     /codon_start=1
                     /product="Ribulose-5-Phosphate-Epimerase"
                     /protein_id="CAA58554.1"
                     /db_xref="GI:609674"
                     /db_xref="SWISS-PROT:P46969"
                     /translation="MVKPIAPSLASDFANLGCCECHKVINAGADWLHIDVMDGHFVY
                     NITLQPIVTSLRSPRPDASNTKKPTAFDFCHMMVENPKWVDFEAKCADQFT
                     FHYEATQDLHLVKLIKSGIKAAAIKPGTSDVLFELAPHLDMALVMTVERPGGQ
                     KFEMDMPKVETLRAPFHLNIQVDGGLGKETIPKAAKAGANVIVAGTSVFTAADPHD
                     VISFMKEEVSRELRSRLDD"
BASE COUNT      259 a 202 c 196 g 254 t
ORIGIN
20.4%; Score 79; DB 45; Length 911;
Best Local Similarity 54.6%; Pred. No. 4.2e-11;
Matches 231; Conservative 0; Mismatches 150; Indels 42; Gaps 2;

QY 5 aaagaaagaaagatgggaatgacacccgaaatagctccttcgactgctcttcgactt 64
Db 104 AAACACACAAAGAAAAATGGTCAACCAATATAGCTCCCAATATCTTGTCTGACTT 163

QY 65 cgccaatgttgcgaggtcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 124
Db 164 CGCCAACTTGGGTTGCGAATGTCATAGCTCATCAACCGCGCGCAGATTGGTTACAT 223

QY 125 ggacatcagtgatggcgactttgttcccaatttaactattggcgctcagttattgaa 184
Db 224 CGATGTCATGGACGGCCATTTTGTTCACCAATTACTCTGGGCCAACCAATTGTTACT 283

QY 185 tttag-----aaagcacacaaaggc 205
Db 284 CCTAGCTGTTCTGTGCGACGCCCTGGCGATGTAGCAACACAGAAAGAGCCACTGC 343

QY 206 atatttgatgtcaccttattgttacaatacctcttgattattgtgaaccttgcaaa 265
Db 344 GTCTCTCGATTGTCACATGATGGTTGAAATCTCTGAAATGGGTGCGACGATTTGCTAA 403

QY 266 agctggtgcttctggttttacatttcacgtagagacatacaaaagataaactggaa 325
Db 404 ATGTGGTGTGCTACCAATTTAGCTTCCACTAGGAGGCCACACAGACCTTTGCATTTAGT 463

QY 326 tatccaaagaatcaactcacatgcatgattcctggtgtagcattaaagcctgggacccc 385
Db 464 TA---AGTTGATTAAAGCTTAAGGCATCAAGCTGATGCGCCATCAACACCTGTACTT 520

QY 386 cgt 388
Db 521 TGT 523

RESULT 8
SCYJL121C/c
LOCUS      SCYJL121C      1328 bp      DNA      PLN      11-AUG-1997
DEFINITION S.cerevisiae chromosome X reading frame ORF YJL121c.
ACCESSION  Z49396 Y13136
VERSION    Z49396.1 GI:1008312
KEYWORDS
SOURCE     baker's yeast.
ORGANISM  Saccharomyces cerevisiae
           Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
           Saccharomycetaceae; Saccharomycetes.
REFERENCE  1 (bases 1 to 1328)

```



```

/translation="LQDIINEFGGAMPQTFGVPVEIIVRGIKMGVYKVNIDTCMRM
RGQFRIIAONKAEFPDRKFLKAPAMDMDRLCKARLEAFGTAGHASKIKVIPMDMAK
RYASGSLAPKTN"
351. .1740
/gene="cbbm"
RBS
351. .354
/gene="cbbm"
CDS
364. .1740
/gene="cbbm"
/EC_number="4.1.1.39"
/notes="rbpL"
/codon_start=1
/transl_table=11
/function="carbon dioxide fixation"
/product="Form II ribulose-1,5-bisphosphate
carboxylase/oxygenase"
/protein_id="AAB82048.1"
/db_xref="GI:727418"
/translation="MDQSNRYARLDLKEADLIAGRHVLCAYIMPKAGYGLLETAH
FAEASTGTNVSTDDTFRGVDALVEIDPEKEIMKIAYPVLEFDRNIIDGGAMIC
SELTUIGNQMGDVEYAKMHDFVPPCYLRLEDPGSMNIADMWRVLPVVDGMV
VLTITKPLGLRPKPFADACEFWLGGDFIKNDPQGNOTFAPLKETIRLVADAMKRA
QDETGEALFSANITADDHYEMVARGEVILETFGENADHVAFLVDGYVTGPAITPAR
RSFPQFLHYHRAGHCYAVTSQSMRGVTAFLVLSKRLQASGIHTGTMGYKMEGDA
SDKIMAYMLNDEAAGPPYHODNLGMKATPILISGGMNALRLPGFDNLGHSNVIQTS
GGGAFGLDGTAGAKSLRQSCDAWKAGVDIVTYAKSHRELARAFESFPNDADKLYPG
WRVALGVN"
1755. .1776
/rpt_type=inverted
repeat_unit
1786. .2483
/gene="cbbg"
gene
1786. .1790
/gene="cbbg"
RBS
1797. .2483
/gene="cbbg"
CDS
/EC_number="5.1.3.1"
/notes="ribulose-phosphate epimerase; cfxE"
/codon_start=1
/transl_table=11
/product="pentose-5-phosphate-3-epimerase"
/protein_id="AAB82049.1"
/db_xref="GI:2564973"
/translation="MTFDRSIKTIAPSLTADAFDFGREIQAIIEAQADWNHVDVMDGH
FVPLNLTGPPAVAFKRVKTYMDVHLMISPVDAIDAYAQAGDVLTAHVEAGPHIH
RLQAIKRAQKHSQVAINPGTAPAEIEHVDIADVVCVMTVNFPGQYQYIDMTAKVR
KLRMIGDRPVIIEIDGMDPVTAPLMAAAGADVFAVAGSVFKGSGSVTPVYVGNIA
AIRAAQAAL"
2492. .3043
/notes="hypothetical protein; similar to A. brasiliense f1ca
gene, encoded by GenBank Accession Number Y12363,
Synecocystis sp. ORP_ID:s111289, encoded by GenBank
Accession Number D90904, M. tuberculosis MTCY04C12.12 orf,
encoded by GenBank Accession Number Z81360, and M. leprae
hypothetical protein MLCB2052.40, encoded by GenBank
Accession Number Z98604"
/codon_start=1
/transl_table=11
/protein_id="AAB82050.1"
/db_xref="GI:2564974"
/translation="MAVSPVPCDFGNKAPDFRLPDFSGRFSILDIRGPKGLLIIMFIC
NHCYPVQAILDLRQDAEALMDLIGVAAISSNDVASYEDFSFANMARLAABERGFRFP
YLDETQAVAKAYDAICTPDPFGFNAECQLYNGRLDASGRTPGPDARRELYEAMLS
VAETGGQPRQIIASMGCSIKWE"
3043. .3702
/gene="cbbz"
CDS
3043. .3702
/gene="cbbz"
/EC_number="3.1.3.18"
/codon_start=1
/transl_table=11
/product="phosphoglycolate phosphatase"
/protein_id="AAB82051.1"
/db_xref="GI:2564975"

```

```

/translation="KAPALIFDLDTGLSDISAPATHKVSNDVLRAGYAPLGLDQIRSF
VQGAHLVRCLLTITAGDEPEGLFDAYADLVSRYETDVEGNTLYPGVITPQALRLRE
MGCPIAITNPKYKPAALAHVGLTDFQVLVIGDGLTPKPNPMEVNRVLRRLR
HALYIGDSEIDAQTAQNAGLPVIYTEGYRTPKLDALPHAAKFHDFSALPGIVEGWTW
S"
3682. .3685
/gene="cbbv"
RBS
3682. .4376
/gene="cbbv"
gene
3693. .4376
/gene="cbbv"
CDS
3693. .4376
/gene="cbbv"
/codon_start=1
/transl_table=11
/product="Cbbv"
/protein_id="AAB82052.1"
/db_xref="GI:2564976"
/translation="MELKALIFSDVDGTLAETEEVHROAFNETFAAOGLDVWWSKEDYR
TLIRTTGCKERMAKHRENILSGSPSDAKIADLHKAKTQRYVEITIASGOVGLLPVGAELI
DRAKASGLRLAIAITTTTANYDALIAITFSKPAGDIFEVIAAGDEVAAQKPAQPVYLR
ALOGGLPAPACLAPEFDSRAGLSARAAGLRVLTPTSEYTRGDDFSAADWRIPDLASA
ATQAIPELP"
4437. .4877
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="AAB82053.1"
/db_xref="GI:2564977"
/translation="MDLRTLTTPDFAVAPQIALDDIAEAARLGFRTLINRNPDEVESG
LNAALQAAAAAGLSVYHLPYYPGELTPDLIAGFEAAALAADAPVLAWCRSGTFRSHL
WALSOAGORPLEEIVQTAQNAGYDLSALVPMRLAKAADSAGVIV"
5054. .5806
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="AAB82054.1"
/db_xref="GI:2564978"
/translation="MLDPLEKLLVADDIATVSNHCAQMAKFGDLRIIVGTFRFSATQ
NLGSLDDTLILSNHDPAYLNSFLGAGLFSAPMWKVAIDSGACSWRMAERNAAGEL
TEGERKVIELNRRFQVLGYSIGFPDTSARAKAIGLCAQPGLTQADVDIADHDEGRD
LRVLSIMHLKISALPSTTRRPLTARQREVLEWVGDKTSADIAQILGLTTATVEKH
LRLAREALEVETTAQALLKSMQKLFITSAP"
BASE COUNT 1123 a 1908 c 1857 g 1072 t
ORIGIN
Query Match 20.3%; Score 78.6; DB 2; Length 5960;
Best Local Similarity 55.8%; Pred. No. 6e-11;
Matches 150; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
QY 33 aaatatagctctctcgatctcttcgcaattcgccaatggcttcgaggtcagcgc 92
|| || || || || || || || || || || || || || || || || || || || ||
Db 1818 AAGATCGCCCTCGATCCTTCGCGGACACTCGCCGCTGAAATCCAGGCC 1777
|| || || || || || || || || || || || || || || || || || || || ||
QY 93 atgtccacttcgcgccgattggtccacatggacatcatgcatggtgcatattgttcccc 152
|| || || || || || || || || || || || || || || || || || || || ||
Db 1878 ATCGAGGCGCAGGCGGATGGTGCATCTCGACCTGATGATGGCCATTCGTGCGCG 1937
|| || || || || || || || || || || || || || || || || || || || ||
QY 153 aatttaacatattgcygtccagttattgaaagtttgaaagacacacaaaggcatattg 212
|| || || || || || || || || || || || || || || || || || || || ||
Db 1938 AACCTCACCTTTGTGCGCGCGCTCGCGGCTTCGCGAAGCATGTGAAAACCGTGATG 1997
|| || || || || || || || || || || || || || || || || || || || ||
QY 213 gatigtcaccttattggttacaatacctcttgattattgaaaccttggcaagctggt 272
|| || || || || || || || || || || || || || || || || || || || ||
Db 1998 GACGTGCATCTCATGCTCGCGGTCGATTCCTATATATCGACGCTTATGCGAGCGGGC 2057
|| || || || || || || || || || || || || || || || || || || || ||
QY 273 gctcttggttttacatttcacgtagagac 301
|| || || || || || || || || || || || || || || || || || || || ||
Db 2058 GCCGATGTGCTGACCGGCCCATGTGCGAGGC 2086
|| || || || || || || || || || || || || || || || || || || || ||
RESULT 10
AE002306
LOCUS AE002306 12939 bp DNA BCT 26-MAY-2000

```


Chlamydia muridarum, section 37 of 85 of the complete genome.
AE002306 AE002160
VERSION AE002306.2 GI:8163215
KEYWORDS
SOURCE Chlamydia muridarum.
ORGANISM Chlamydia muridarum.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
1 (bases 1 to 12939)
Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F.,
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,
Bowman, C., Dodson, R., Winn, M., Nelson, W., DeBoy, R., Kolonay, J.,
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
20150255
REFERENCE
2 (bases 1 to 12939)
Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F.,
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,
Bowman, C., Dodson, R., Winn, M., Nelson, W., DeBoy, R., Kolonay, J.,
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7190428.
COMMENT
FEATURES
source
Location/Qualifiers
1. .12939
/organism="Chlamydia muridarum"
/db_xref="taxon:83560"
/note="synonym: Chlamydia trachomatis MoPn"
complement(115..1941)
/gene="TC0388"
complement(115..1941)
/gene="TC0388"
/note="similar to GB:232522 PID:510140 SP:P54124
PID:1771208; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligodehydrogenase F"
/protein_id="AAF39245.1"
/db_xref="GI:7190429"
/translation="MTDTTSQTLRSRDKVPSDCWDTKSLYASREAWKKDLNLCG
KEAPFPHNEQFHEOPSSLYKLLTEVFSEIRLEKLYVYVAHLTHDDISNQATA
DLKINYLITSESEISWIOALIALPQKISELITSPLOTVHFVLYKIPRLAPHTG
TSREKILASSPALVAYKTFSSLDSDIPPEGEAVDSGSKSHPLSHALSSLYMQSPD
RELKSAQKQORYHGYRLSLANLNGIKQALHFNKAKHNTKSCLEAALFONNISTS
VVTTLTKQHTLITKYFQKQKALGDPDFHVDYAPIAASEFARHYSEEAVALS
ICDLPPLGDNVYALRKGLTSEGWDYENQNRKSGYSSGCDYKPYLLNVTGTL
YDVSVVAHEGSHMSHFLSHKHQYHEAQYPIFLAETASTLNETLMEPLKNAPSKE
EKTAILRSLEDTIFATLFRQTLFADPELAHSAEEOGIPLEETFSQSYKQLOHIFYG
DCTFPELSIENWAPHYPIFYNYVYATGIIASLCFSEKILSKREAGAEALYLFQ
SGVDFPIELKSGLDMTSSAPMLAAFSYIERKLDLNTLL"
2237..4840
/gene="TC0389"
/note="TC0389"
/db_xref="GI:7190430"
/translation="MEKFSVASEALEKAPELAKDAKSHSVTENVHLLKLLONPGSLF
CIVIKOVHNGLLTNSVDALHREKTVVGEAALPKPSSLOQLLINAQEEARSMGDE
YLSGDHLLALFATKPEPFASWRKTVKTSPEALKELIILKQSRMDSFAENLKL
EKYCKNTIILAREKGLDPVIGRDEIRITQVLSRRKNNPMLIGEPGVKGAIAEGL
ALRTVQGVPESLKDLHYLDGMAIAGAKYRFEERKLYKLVKGEVSEACILFI
DEVTLLVCAGATGDGMAANLLKPALARGTLHCIGATTLNEYOKYIEKDAALERRFPQ
IFVTEPSLEDAVFLRLGKREKYEIFHGVRTGTEGALNAAVLSYRITDRFLPDKAIDL
IDEASLIRMQISGLPLPIDEKERELISALIVKQEAIKREQAPAYQEAEMQKADRY

KEELAALRLRWDEEKGGLIAGLKEKKKNLENLAKFAEEAERTADYNKVAELRSLIPSL
EERLAEALNORDGRLLQEEVDERLIAQVNVANWTGIPQVKMLGESEKLLVLESFL
EERVQGPFAITAVSDSISARVGLSDPQPLGVFLGPTGVGTAKALAEALLFN
KEEAMFDMETWEKHSVKLLGSPGVGVEEGGSLSEALRRRPSYVVLDELEKA
DKEVFNILQIFDDGILTDKSKKKNVCKNALFIMTNSIQSELCAKKGATVDEKA
VLSVAPALKNFYSPEFINRIDFLVPLTTEDIVKVIQIMNRVALRLLERRISLT
WDDSVVFLSEQYDGAFGARPLKRLIQKRVVVTMLSKALLKGDIKSGMSVELTMAKDV
VFKTKTNSVV"
4850..6325
/gene="TC0390"
/note="TC0390"
/db_xref="GI:7190431"
/translation="MMKHWIRIIGISLSPVLLALGEXTPAFFLERNRNFNIDPY
CLISACFTSNODPLSAERLVLFQLOSDLLVFAKYLSSKNAYLFSDAEERIF
AKTALPRVSLGNCNRADDLAKVLAEAGSTAKQKARYSLYLDVALRAYVBRARLV
NATCEVASQIDIASIEAINTILFEKSGVRVPSKEMFESFSELAATDSKFGVCLGS
VVLQVAQRNLNLSLEAVTPPGHIYLRKDVNIETSSGRHIPTKYECEIPESOLR
VROEELIGLTFMNRGAFFLOKGEFLOASIAEYQAKYLADEQISDLGITYILLGK
KEGEDILKSSERTKRKSAIHDFERNISPEILAVLPADSGVTVQETLDYHKKLLVIV
EKPRSGALRLRAATALEGLVKGEQLLEKSEVEDAPDDLFLRLQLKILCNRHDT
RAKHFDQAKKILAEEGMLSESSYTLKLEKKLALVSPH"
6538..6912
/gene="TC0391"
/note="TC0391"
/db_xref="GI:7190432"
/translation="MTKICSTSQGLAACRIERINLEKSLNPRNAAPTTRKLTJVALA
AISAITFLVIAGLLFVGVLCPPVLIASALLGLGAFGLGTVVGGLLTVEAAAKDRL
YRSQVLSNNLCKYKTAGVIRNG"
7009..7440
/gene="TC0392"
/note="TC0392"
/db_xref="GI:7190433"
/translation="MECIRFCCKDLCKLVNPNVCSLKKKTCERKQVQVAGCIGVI
CSMVCLALGIAAGCTGLIGRAFTVSLGVALLGLVLFMSIYDVLERHGLGCPKMT
REPAPEPAPEPAPELSVKTGQDQEDCAIVGCCRK"
7452..7736
/gene="TC0393"
/note="TC0393"
/db_xref="GI:7190434"
/translation="MIQSIERESTFIESHRPCDLSKHKAKAFAFGIAKLIALVALIL
NGALPALSIVIAVCVGSSTPASPVLGAAATLASFLCAARVFLITKDRGW"
7771..8271
/gene="TC0394"
/note="TC0394"
/db_xref="GI:7190435"
/translation="MISQIERESTFIESHRPCDLSKHKAKAFAFGIAKLIALVALIL
NGALPALSIVIAVCVGSSTPASPVLGAAATLASFLCAARVFLITKDRGW"
7771..8271
/gene="TC0394"
/note="TC0394"


```

/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39251.1"
/db_xref="GI:7190435"
/translation="MDGRMVCCGKILLEGQOTLPTVERBSSICAIVVKKTALLAVSL
FAVLTAGSLAVLSGLVFSGTASPLPYLLIFAAALFGVCASIVLVKCISSVVQDVKK
KEKAPESIPNSDLSSSSSSSELEKLVDSSSLGSLISIGDCVDEAFGRNRELRAIR
ARANS"
8207..8383
/gene="TC0395"
CDS
8207..8383
/gene="TC0395"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39252.1"
/db_xref="GI:7190436"
/translation="MKHLGLGTMSCGLGLEQIAVNERRGSEPLLFYHFRSILICPCN
KVLISQGFFFKVA"
8207..8383
/gene="TC0396"
CDS
8207..8383
/gene="TC0396"
/note="similar to GP:3172545; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="inclusion membrane localised protein Inca"
/protein_id="AAF39253.1"
/db_xref="GI:7190437"
/translation="WNSPLVEMPLSCYPTHTSSTACTKRSSLYKPSLIETVQVRA
AFVSLAISTIGLAILGHAIGLAPQALVLIATFIISLGNALYIACKTAPLRKLV
EDQEVASKEVNFLLSKQDFATTSKDLSDVLDHNLQDFQSSHOFG
EDLEEDKNSAELRQIFSOETVQSLKSTILSKBEIKVPLTEVRLAENKEDLL
KIVQDLQIRDKLRABINNLQSKATLSQIASQIEENEKLYANITKALSQSYSDSSA
"
9703..10389
/gene="TC0397"
CDS
9703..10389
/gene="TC0397"
/note="similar to PID:606320 PID:1789788; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribulose-phosphate 3-epimerase"
/protein_id="AAF39254.1"
/db_xref="GI:7190438"
/translation="MKKQGLIAPISIMGADLACLGDAAARITEESGANLIHIDVMDGHF
"
Query Match 20.0%; Score 77.6; DB 1; Length 12939;
Best Local Similarity 53.4%; Pred. No. 1.2e-10;
Matches 186; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

Qy 35 aatagctcttcgatgctcttcctccgactcgcacattgcttcgagcgatcgagcgc 94
Db 9723 AATTGCTCCATCATATATGCGGCCACACTTGCTTCCTAGGGATGACGACGAAATAT 9782
Qy 95 gctccacttgccgagcgtgctccacatggacatcatggtatggcgtttttgtcccaa 154
Db 9783 AGAAGAGTCGCGGCAAACTTATTATATAGATGTCATGATGGCATTTTGTTCGAA 9842
Qy 155 tttaacctattggcgtcccgattatgaagtgtgagaagacacacaaaggcatattggga 214
Db 9843 TATTACCTTGGTCCCGGTATTATGCCCAATCAATCGGTCAACAGATCTATTCTTGA 9902
Qy 215 ttgtccacctattggttacaaacctcttattgattgtaaccttggcaaaagctggtgc 274
Db 9903 AGTTTCATGCTATGATTATACGCCGCTTGTGAATTTGTAGAGGCTTTGTAAAGCTGGAGC 9962

```

```

QY 275 ttctggttttcatcttcacgtagacacatcaaaagatactggaaagaaacttatccaaag 334
Db 9963 AGATCGTATTATTGCGACTTTGAGGCAGC---AGAAACCTTAAGAGAAATCTTTGATTA 10019
QY 335 atcaagtcacatggcgtgattcctggtgtagcattaaagcctgggac 382
Db 10020 TATTCGAAATGTGGAGTCAGGCAGGAGGATCGCTTTTCTCCAGAGAC 10067

RESULT 11
AE002217 10150 bp DNA BCT 30-MAY-2000
LOCUS Chlamydomophila pneumoniae AR39, section 45 of 94 of the complete
DEFINITION genome.
ACCESSION AE002217 AE002161
VERSION AE002217.2 GI:8163462
KEYWORDS
SOURCE Chlamydomophila pneumoniae AR39.
ORGANISM Chlamydomophila pneumoniae AR39
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE 1 (bases 1 to 10150)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
2 (bases 1 to 10150)
REFERENCE Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7189493.
FEATURES
Location/Qualifiers
1..10150
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(122..1294)
/gene="CP0581"
complement(122..1294)
/gene="CP0581"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38399.1"
/db_xref="GI:7189494"
/translation="MSPVNTPTSPAPNIPTTGGITTKPRSSPIEKVIVKVI
LFAIATGALGTILGLSCALTPGIATLVIFFVSMVLLGLLKLKDSISGSEERRLE
EVSRTSENORLTVTITTEVTKDLKAKDOLLEAFRNENGNLKTFAEDLEQV
SKLSOLEALERINQLOANAGDAQEISSLKLIQSWDSKVVEQINTSQALKVLGG
QEWQEAOTHYKMQEQIQALQAEILGHNSQALQSVENLQVQDALFRVYVGLLE
SENKUSQACSALEQIEIQAQHEHTSLQORIDAMLAQEQNLAEQVTALEKMQEAKAE
SEFTIACVDRDTFGRRTPPTTTPVVEGDEGEGTTPPVSPSPVDRATGDGQ"
complement(1349..1456)
/gene="CP0582"
complement(1349..1456)
/gene="CP0582"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11

```

[illegible]

QY 35 aaatagctcttcgatgctctcttcgcgacttcgcgcaattggcttcgcgaggtcagcgcat 94


```

/protein_id="BAA98211.1"
/db_xref="GI:8978374"
/translation="MLGKIIRGLSLVILCALNVLGIGITHINKNIITAKLCGGVSTP
ACQITIIIGIAGVICLLSFCPCSKSRHSHGSDSCSGGCHSHSDKN"
573..878
/gene="gatC"
CDS
573..878
/gene="gatC"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (C subunit)"
/protein_id="BAA98212.1"
/db_xref="GI:8978375"
/translation="NEOHLDEEITLLAKASALQSLSEELIQEVQTSLSAVITSKKEA
LAIEDDADSCSLFHWVNVEDLREDSVTSDFNREEFLRNVPESLGLVAVPAVK"
895..2373
/gene="gatA"
CDS
895..2373
/gene="gatA"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (A subunit)"
/protein_id="BAA98213.1"
/db_xref="GI:8978376"
/translation="NYRYSALELAKAVTLGELTATGVTOHFFHRIIEAEGOVGAFISL
CKEALQEALELDKKRSRGEPLGKLAGVPGIKONIHVTKLTTCASRVLENYQPPD
ATVVERIKKEDGILLKLMDEFANGSTLTLSAHPHTNPWDLRSKVPGGSGGSAAV
SARFCVALGSTGSGIRQPAFCGVGKPSYSAVSRYGLVAFASLDQIGPLANTV
EVALMMDVFGSDPKDATSEFFDSFMSKLTSEVPKVGIVPTFLEGLRDDIRENF
FSLAIFEGEHTLVDELILSHAVSIYIILASAEATNLARFDGVRYGVRSPQAH
ISQYDLSSGEGFGEVMRRILLGNVLNVAERONYKKATAKRAKIVKAPRTAFEC
ETLAMPVCSGPAFEGEITLDPTVLQDIYVANNLVLPAIVLPAIVGFSKRGELPLGQ
IIGQOGQOQVCQVCYSQEAHQKQLFSKRYAKSVVLGGQS"
2370..3836
/gene="gatB"
CDS
2370..3836
/gene="gatB"
/codon_start=1
/transl_table=11
/product="Glu tRNA Gln amidotransferase (B subunit)"
/protein_id="BAA98214.1"
/db_xref="GI:8978377"
/translation="MSAVYADWESVIGLEHVHVELNTASKLFPSSALNRFGEPPNTNIST
VCTGLPGSLPVLNQSAREKAVLECAVEGEISLRSRDRKSYFYFDPSPRNFQITQFHE
PIIRGRKAIYQGEERIFELAQTHIEDDAGMLKHFGFAGVDYNRAGVPLIEIVSKP
CMCPEDAVATSLVSLDYGISDCNMEEGSIFDYNVSRPKGSPPELRNKVEIKN
MNSFAQAALAEAKQRIQIDELQPNKPKLVIPTAATYRWDPEKKTKVLMRLKESAE
DYKPEPEPLTLQLTESIYERIKTLPELPYKHYRYIQEYGLSEDIATLISDKNI
ATFPEVACKDNFSLNWSYVEFGGRCKTLGKLPSSGIFPEGVAQLYNAIDQGV
TGKIAKETADLMMESPCKNPEILKERPELLPMDSDEGELQKIIAEVVLAMPESIVDYK
NGTKALGFLVQGMKRTAGKAPPRVNNELLILLBQK"
4127..6895
/gene="pmp_1"
CDS
4127..6895
/gene="pmp_1"
/codon_start=1
/transl_table=11
/product="polymorphic outer membrane protein G family"
/protein_id="BAA98215.1"
/db_xref="GI:8978378"
/translation="MRFSLCGFPLVFSFTLLSVFDTLSATTISLTIPEDSFHGSQNA
ERSYVQAGDVYSLTGQVYSINVDNSALNKAFCNVTSGSVTFAGNHGLIFNNISST
TKEGALCCQDPQATARSPGFSTLSFQSPGDIDKEQGLYSKNAIMLLNNYVYRFEQN
QSKTGGAGISGANTIVIGNYDSVSFYQNAATFGGAIHSSGVLQIYVNAQAEIRFAQNTA
KNGSGGLDGDIDQNAVYLFRENEALTATCKGGAACVCLPTSGSSTPVPIVTS
DNKQIVFRNHSIMGGAIYARKLISGGPTLFINNISYANSONLGAATAIDTGGPI
SLSAEKGITFOGNKTSPLPLNGIHLQNAKFLQARNGYSEIFYDPIITSEADGTSQ
LINQDPKNKYEYCTILFSGEKSANDPRPKSTIPQNVNLISAGILVYKGAETVYSK
FTQSPGSHLVLDLTGLASKEDIAITGLAIDIDSLSSTAAVIKANTANKOLSVTD
SIELISPGNAYDLRMRNSQTFPLSLEPGAGSVTVTAGDLPVSPHYQFGSNMKL
AWTGTGNKVGFFWDKINYPKPEKGNLVPNILWGNADVRSIMQVQETHASSLOTD
RGLWDIGTGNFPHVSASEDNIRYRHNSGGYVLSVNNETTPKHVTSMAFSQLFSRDKDY

```

```

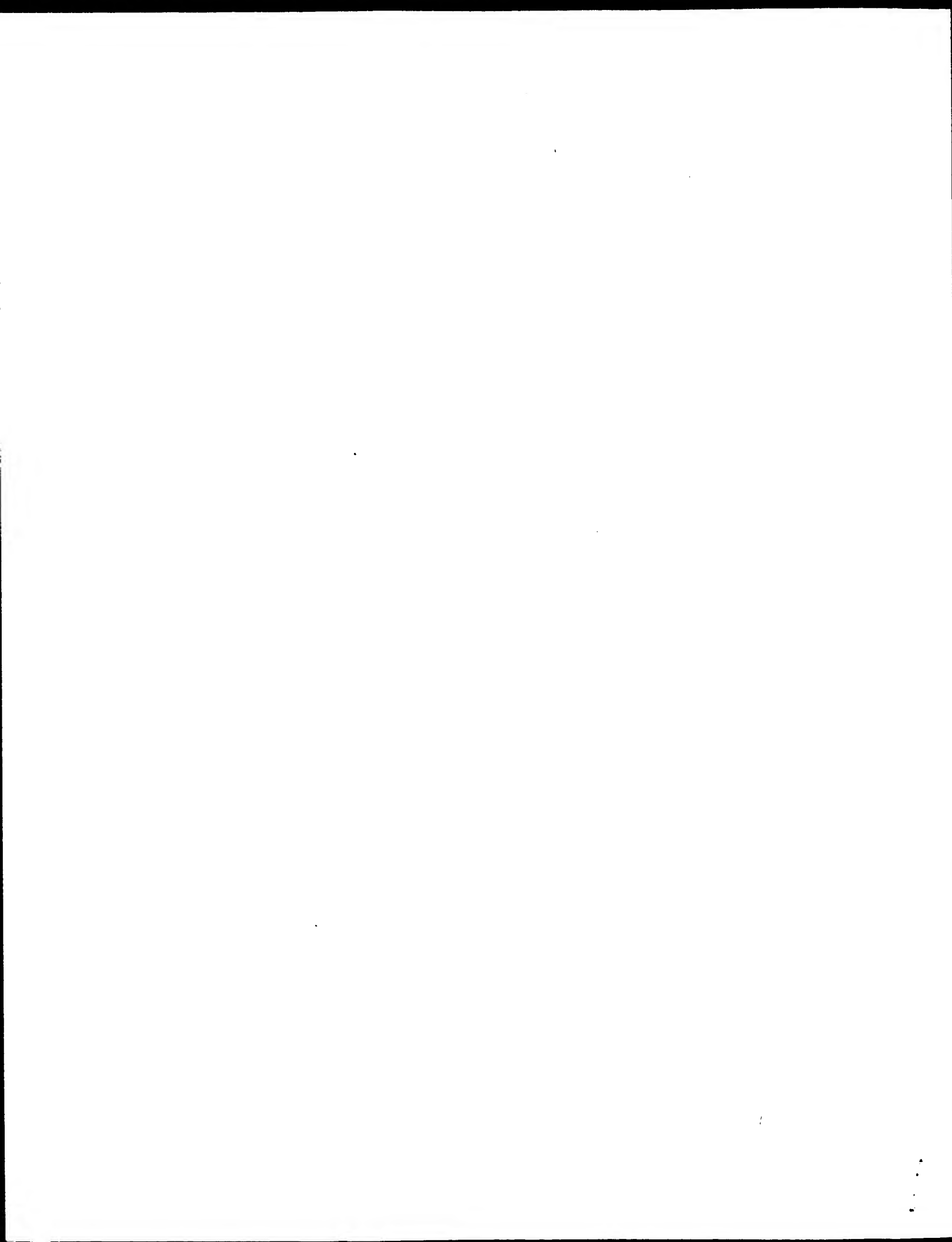
AVSNNEVRYMLGSLYQYTTSLGNIFRYASRNPNVNVNNGILSRFLQNPIMIFHLCAV
GHATNDKTDYANFPMVKNRNNCWAIECGSGMPLLVFENGRLFOCAIPFMKLQLVY
AYQGFKEFTAGRRFNGSLTSSIVPLGIRFEKLALSOQVLYDFSPSYIPDIFRKDP
SCEAALVISGDSWLVPAAHVSRHAFVSGTGRYHFNNDYTELLCRGSEICRPHARNYNI
NCGSKFRF"
complement(7138..7293)
/gene="CP10006"
complement(7138..7293)
/gene="CP10006"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98216.1"
/db_xref="GI:8978379"
/translation="MQEPLRSALLERLEWLVLLGVSPETTRSTPEKQANQLPKDSR
NRTLESL"
7605..10499
/gene="CP10007"
7605..10499
/gene="CP10007"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98217.1"
/db_xref="GI:8978380"
/translation="MRKRLIALVIALSILIAAGVVLITVAIPGLSSVSSIPAGMG
ACALGCVMLAGIDLVLKKREVPILVASVTTTGTGTPSPRSGISIGADSTIRSLPTVL
LDGHPGQMRKRLIALVILVIFSLIIASGVLLITVAIGLSSVSPARMGACALGC
VMLAGDLVLLKREVPILVASVTTTGTGTPSPRSGISIGADSTIRSLTYLDEGHP
OSMRKRLIALVILVIFSLIIASGVLLITVAIPGLSSVSSIPARMGACALGCVMLAG
IDVLKKREVPILVASVTTTGTGTPSPRSGISIGADSTIRSLTYLDEGHP
VESHLENNKSLPYDGHGLEKTKHQIRVRSRSLKAMVPEFLDIRIFEEFEFFLSAR
KRILDTLTVVERKIITQELERNLKAFLYQDSIFKKIIDNFKLAKWLKILHGNPFE
ICRTIIFENHEHGMVAKSLHKNVLLKVIYRSLOKSYRDIQSGSAAKWLHGNPFE
ISRDQWQKAHLKHQESLYTQADRDLTQSSRENNKELEKEQETISSWERVKKFEI
ERYQRIQAIQKLYPNILREETTGQETVPTVQGTASSDLTBIILGRIEVSREDN
ONQESCVKLSRHEVSEVMEVQKQYKPKFQDQMSGLERFTEHIELEVLQKQYS
KHLSTFKVNNKKEVQYAKFRLKQSDSLGILAQTESAELTQBELPLATRGALE
KAVFGKSLCALKAKAPYEPDPRFQSDTOLRALTLRLOEAKASLESEIKRFSNLE
NDIAEERRLLKESQTFERAGLGLREIAVESTYDLRSLTNTWEGTPESEKYVFSYL
NYNEEKRAKTRILVEMTQYRDFKMALEAMQNEEALLQEBLSIQAPSE"
10975..111688
/gene="CP10008"
10975..111688
/gene="CP10008"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98218.1"
/db_xref="GI:8978381"
/translation="MISGLLFLVLRREVPTVSEETPRGVSVPSEPALEAKQKPE
TKKILDRPKELQDQTYIQEVEACLERLQPKYEDRGLLTAKEKLRVDPVDEKDM
SEFLDQRLVNEEAYVYEHQCDPLENIAIEYSSQELRDYICAGVCGYLPSGDARRDR
LKRSVREMDRMFRVTKWSWEASVMLDHSYIGVARELFRKAVGVLEESYKILFKSYRD
AFYCEKAKIQDRGREFKL"
11815..13122
/gene="CP10009"
11815..13122
/gene="CP10009"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98219.1"
/db_xref="GI:8978382"
/translation="MKLDYADKFLIRKEEKMERHELPHATWPKASCHAYAKAKAA
FEKERNENQKRVKQVKKWLSKGLAEFRQESRRERELQTLIPYVSVEERVLER
QRTKYNLENLADIIEKKYHHCVRQEHYKVENKEAYRENGEKLVSAAEVSECLQ
RIEDCLETWPKLITKAESEVFEMKFDATKGNKLVSLDVTNRLIEICEAEMIEFRIE
ETIMTLMWMLPLLFEMKQYNSCKEMAKLAKVEPOCKSPYTRSSQRLRLRN
QDLOATYTNQERLQGFDSLEKVRTRCDRLHOMKHFVEQGLNFTNEELLGAEFL
TOARLDLATVPYMEFYLOYHNILKREKVSQWMAKTEREIRIQAFQGVYKEDLIAED
TILKEEDYLLRDDLLRBERKNRRLICNKLAAQAQKVKGF"

```


JOURNAL	Submitted (06-JUN-1999)	Life Sciences Division, Oak Ridge National Laboratory, 1060 Commerce Park, Oak Ridge, TN 37831, USA
FEATURES	Location/Qualifiers	
source	1..993	
	/organism="Expression vector pFL505"	
	/db_xref="taxon:78121"	
	/lab_host="Escherichia coli"	
	/note="synthetic expression cassette for Spinacia oleracea chloroplast ribulose-phosphate 3-epimerase transit form"	
gene	1..993	
	/gene="RPE"	
promoter	1..97	
	/gene="RPE"	
	/note="synthetic tac (trp-lac) promoter"	
CDS	98..955	
	/gene="RPE"	
	/EC_number="5.1.3.1"	
	/note="pentose-5-phosphate 3-epimerase"	
	/codon_start=1	
	/transl_table=11	
	/product="ribulose-phosphate 3-epimerase transit form"	
	/protein_id="AAC24709.1"	
	/db_xref="GI:3264790"	
	/translation="MGAAALCSTLQSQINGFCGGINIRKLQPSSTSPNSLFTTRKY QTLVKATSRVDKFSKSDIIVSPILSANFAKLEQKVAELAGCDWIHVDVMDGRFVP NITIGLVADALRPVLDPLDLHLMIVEPEQRPDPIKAGADIVSVHCSELASTIHLHR TYNQIKSLGAKAGVNLPGPTLTIEYLDVDDVLIMSNVPGFGSQFTIESQVKKIS DKRMCKVEQGNVPWIEVDGGVTPANAYKIEAGANALVAGSAVFAGKADYAEAIKIGIA SKRRPFPVAV"	
BASE COUNT	268 a 202 c 237 g 286 t	
ORIGIN		

Query Match	19.4%;	Score 75.4;	DB 91;	Length 993;
Best Local Similarity	53.0%;	Pred. No. 4e-10;		
Matches	185;	Conservative	0;	Mismatches 161; Indels 3; Gaps 1;
QY	40	ctcttcgatgctcttcgcgaccttcgccaatttgcttcgaggtcagcgcatgctcc	99	
Db	291	CTCCATCTATTCTCTGCTAACTTTGCAAAAGTTGGGTGAGCAGGTAAAGACGATTGAAT	350	
QY	100	acttcggcgccgattggctccacatggacatcatgatggcgatttggccccaatttaa	159	
Db	351	TGCGCTGGTTGATGATGCATGTTTGATGATGATGGTCTGTTTGTGCCAAATATTA	410	
QY	160	ctattggcgctccagttattgaaagtgttgagaaagcacaaaggcatttgggattgtc	219	
Db	411	CTATTGGAGCCTCTGGTGGTTGATGCTCTGGACCTGTGACAGATCTCCACATAGATGTGC	470	
QY	220	accttatggttacaaatcctcttattattgttgaaacctgttgcaaaagctggtgctcttg	279	
Db	471	ATCTGATGATTGTAGAACGACAGACGGAGTCCCGACTTCATCAAAAGCTGGTGTGATA	530	
QY	280	gttttacatttcacgttagag---acatcaaaagataactggaaagaacttatccaaagaa	336	
Db	531	TAGTGAGTGTCCACTGTGAGCATAGCATCACCATTTCATTTGTCATCGGACCGTAAACCAA	590	
QY	337	tcaagtccatcggcgatgtctctggtgtgagcattaaagcctggagcccc	385	
Db	591	TCAAGAGTTTGGGAGCTAAAGCTGGAGTTGTCTTAAACCTCTGGAACCCC	639	

Search completed: November 4, 2000, 13:36:23
Job time: 17232 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:45:01 ; Search time 320.8 Seconds
(without alignments)
454.355 Million cell updates/sec

Title: US-09-300-482-298
Perfect score: 388
Sequence: 1 ggaagaagaagaagaatg.....ttaaaagcctggagaccccccgt 388

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36 : *
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT : *
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT : *
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT : *
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT : *
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT : *
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT : *
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT : *
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT : *
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT : *
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT : *
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT : *
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT : *
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT : *
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT : *
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT : *
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT : *
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT : *
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT : *
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT : *
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT : *
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
C 1	83.6	21.5	18 V75460	Staphylococcus aur
C 2	77.8	20.1	17082 20 X13166	Enterococcus faeca
C 3	77.6	20.0	12566 19 V52282	Streptococcus pneu
C 4	74.2	19.1	32768 20 X20515	Polynucleotide seq
5	69	17.8	1050 21 X24513	D-ribulose-5-phosp
6	65.8	17.0	1260 21 X24512	D-ribulose-5-phosp
7	65.4	16.9	1234 19 T98644	DNA encoding a S.
8	65.4	16.9	1542 19 T96346	S. pneumoniae deri
C 9	58	14.9	1019 21 A26928	Essential staphylo
10	53.2	13.7	1040 19 V65230	DNA encoding a S.
11	43.6	11.2	459 21 A31245	Plant microsatelli
12	41.8	10.8	11443 19 V52182	Streptococcus pneu

C 13	38.2	9.8	6866	13	Q27406	Human parainfluenz
14	37.6	9.7	439	21	A31286	Plant microsatelli
15	36.4	9.4	573	20	X61720	B. burgdorferi ant
16	36.4	9.4	606	20	X61719	B. burgdorferi ant
C 17	34	8.8	3781	19	V26296	Phosphate starvati
18	33.6	8.7	434	21	A31266	Plant microsatelli
19	33.4	8.6	1222	21	V256773	Human transmembran
20	33.2	8.6	1486	19	V71042	CDNA encoding (iso
21	32.2	8.3	4741	19	V23686	L.lactis MG1363 Ad
C 22	32	8.2	20633	20	X13213	Enterococcus faeca
C 23	31.8	8.2	1380	21	A26697	Candida albicans p
C 24	31.6	8.1	10555	20	X13170	Enterococcus faeca
C 25	31.6	8.1	10749	20	X20257	Borrelia burgdorfe
C 26	31.6	8.1	40352	19	V02032	MAGE-B cluster DNA
27	31.4	8.1	7646	18	V74432	Staphylococcus aur
28	31.2	8.0	1327	19	X14338	H. pylori GHPO 236
29	31.2	8.0	1440	19	V65769	Campylobacter jeju
C 30	31	8.0	1430	16	Q92525	Nicotiana plumbagi
31	30.8	7.9	506	21	V57037	Human secreted pro
C 32	30.6	7.9	506	21	V57037	PBR-associated pro
C 33	30.6	7.9	3150	19	V05542	Actinobacillus ple
34	30.4	7.8	969	20	Z07815	S. aureus nrdf pol
35	30.4	7.8	5897	18	V74831	Staphylococcus aur
36	30.4	7.8	580073	18	T58840	Mycoplasma genital
C 37	30.4	7.8	1230025	20	X91990	Nucleotide sequenc
38	30	7.7	14066	20	X99556	Nucleic acid sequen
39	29.8	7.7	2622	18	T90461	Thermoanaerobacter
40	29.8	7.7	2622	20	Z19861	Thermoanaerobacter
41	29.8	7.7	4617	17	T72853	Rat tripeptidylpep
42	29.8	7.7	4630	20	V84139	Human prothrombina
43	29.8	7.7	5109	17	T72654	Rat brain homoloqu
44	29.8	7.7	5300	18	T90447	Thermoanaerobacter
45	29.8	7.7	5300	20	Z19859	Thermoanaerobacter

ALIGNMENTS

RESULT 1	
V75460/c	
ID V75460 standard; DNA; 916 BP.	
XX	
AC V75460;	
XX	
DT 16-MAR-1999 (first entry)	
XX	
DE Staphylococcus aureus contig SEQ ID #1149.	
XX	
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;	
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;	
KW skin infection; surgical wound infection; scalded skin syndrome;	
KW toxic shock syndrome; ds.	
XX	
OS Staphylococcus aureus.	
XX	
FH Key Location/Qualifiers	
FT misc_feature 1..60	
FT /tag= a	
FT /note= *these bases represent a line of missing text in	
FT the sequence listing in the specification. They	
FT are included to maintain the nucleotide numbering	
FT given in the specification for this DNA sequence*	
FT	
EP786519-A2.	
XX	
PD 30-JUL-1997.	
XX	
PF 07-JAN-1997; 97EP-0100117.	
XX	
PR 05-JAN-1996; 96US-0009861.	
XX	
PA (HUMA-) HUMAN GENOME SCI INC.	
XX	

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
XX
PS Claim 1; Page 1854-1855; 3271pp; English.
XX
XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against *S.aureus* infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the *S.aureus* DNA sequences contained on the
CC computer readable medium.
XX
XX Sequence 916 BP; 273 A; 165 C; 120 G; 291 T; 67 other;
SQ

Query Match 21.5%; Score 83.6; DB 18; Length 916;
Best Local Similarity 53.6%; Pred. No. 4.8e-17;
Matches 192; Conservative 2; Mismatches 161; Indels 3; Gaps 1;

Qy 31 cgaataagcttcctgagctctccgacttcgccaatttggcttcgaggtcagc 90
Db 745 CAAACATATATCCATCATATATATCTGTTGATTTTGGATTACACATGATTAAC 686

Qy 91 gcatgtccacttcggcgaggttggtccacatcggacatcgtggtggtcattgtcc 150
Db 685 GACITGAGAGAGCAGCGCTGACGGAGTTCATTTTGTATGTTGATGGTCAATTTGTG 626

Qy 151 ccaatttaactattggcgctccagttattgaaagttgaaagacacacaaagggcatatt 210
Db 625 CTAATATATCTATTTGGTTTACCAATATATAGATGAGTAAAGAAAGGACACACATTACCTA 566

Qy 211 tggattgtcaccttatggttcaaatcctcttggattgttgaacccttggcaaaagctg 270
Db 565 TAGACGTACATTTGATGATTGAATCCAGAAAGTATATGTCATCATTTGCAGAACATG 506

Qy 271 gtgctcttggttttacattcagtagacatcaacaaagataactggaagaaacttatcc 330
Db 505 GTGCGGATATGATTTCAATTCATCTCGA---ATCAACGCCCTCATATTATCATGCTGCTATTC 449

Qy 331 aaagatacaagtcacatggcatgattcctggtgtagcattaaagcctgggaccccccgt 388
Db 448 AAATGATTAAACATTTAGATAAANAAGCTGGTGTAGTAAATTAATCCCTGGTACMCCMAT 391

RESULT 2
X13166/C
ID X13166 standard; DNA; 17082 BP.
XX
XX X13166;
AC
XX
XX 19-MAR-1999 (first entry)
DT
XX
XX Enterococcus faecalis genome contig SEQ ID NO:229.
DE
XX
XX Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.
XX Enterococcus faecalis.
XX WO9850555-A2.
XX
XX 12-NOV-1998.
PF
XX 04-MAY-1998; 98WO-US08985.
XX
XX 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI
XX Barash SC, Dillon PJ, Kunsch CA;
XX WPI; 1999-045171/04.
DR
XX New isolated *Enterococcus faecalis* polynucleotides and polypeptides
PT - used to develop products for the detection of *Enterococcus* and for
PT use in vaccines for prevention or attenuation of *Enterococcus*
PT infection.
XX
PS Claim 1; Page 1149-1157; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the *Enterococcus faecalis* genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the *Enterococcus faecalis* genome with
CC commercial importance. The products can be used to detect the presence
CC of *Enterococcus faecalis* in samples. They can also be used for
CC diagnosing *Enterococcal* infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of *Enterococcus faecalis*, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences
CC can be used in vaccines to prevent or attenuate an *Enterococcal*
CC infection.
XX
XX Sequence 17082 BP; 4548 A; 3830 C; 3167 G; 5533 T; 4 other;
SQ

Query Match 20.1%; Score 77.8; DB 20; Length 17082;
Best Local Similarity 52.3%; Pred. No. 1.5e-14;
Matches 197; Conservative 0; Mismatches 177; Indels 3; Gaps 1;

Qy 9 aaagaaaagatgggaatgacacccgaaatagctccacttcggcgaggttggctccacatggac 68
Db 7372 AATATAAAGGAGCGAGCGCAAAATGAACATAGCACCATCCATTTTAAAGTCCGATTTTGGC 7313

Qy 69 aatttggcttcgaggtcagcgcatgctccacttcggcgaggttggctccacatggac 128
Db 7312 AATTAGAAAGAGATATCCGTTTAGTAAAGAAATAGGAGCAGATATATTCATGTGTAT 7253

Qy 129 atcatggatggcgatttgcctcccaatttaactattggcggtccagttatgaagtttg 188
Db 7252 GTCATGACGGTCAATTTGTGCCCAACATACCTTAGGACCGAATATCCTTTACGGAT 7193

Qy 189 aaaaagcacacaaggcagcatatttggattgtccacttattgttacaactcctcttattat 248
Db 7192 CGTCCAGTGACAAAACATACCATTTGGATGTTTCATTTTGTATGTTGTACAACTGAAATAT 7133

Qy 249 gttgaacccttggcaaaagctggttcttctggttttaccatttcaacgtagagacatacaaa 308
Db 7132 ATCAATGCTTTTGGGAAGCGCGTGGGATATATTTACCGTCCATCAAGAAGCAACACC 7073

Qy 309 gataactggaaaacttatccaaagaactcaagtcacatggcgatgattcctctgtgtagca 368
Db 7072 CATA---TTCATCGTGCCTTACAAATGATTAAATAATGCGGTGTGAAGCTGCGGTANCC 7016

DR WPI; 2000-097543/08.
 XX P-PSDB; Y54150.
 PT Nucleic acid encoding D-ribulose-5-phosphate-3-epimerase is used to
 PT increase resistance to root nematodes -
 XX
 PS Claim 1; Page 44-45; 62pp; French.
 XX
 CC The present sequence encodes an Arabidopsis D-ribulose-5-phosphate-
 CC 3-epimerase (RPE) nematode response protein. The enzyme is activated
 CC during the initial stages of nematode infestation of roots and is
 CC required for formation of the giant feeder cells on which the nematodes
 CC depend. Interfering with expression of the protein will thus inhibit
 CC nematode infestation. The RPE polynucleotide is used to produce
 CC transgenic plants (specifically potato, tomato, beet, rape and rice)
 CC that are resistant (or at least less susceptible) to nematode
 CC infestation. This is achieved by either expressing antisense sequences
 CC in the plant; or by using the promoter of the RPE gene to express an
 CC anti-nematode protein. Inhibition of the RPE protein or gene enables
 CC plant to mount an immediate response to the invasion of the nematode,
 CC consequently limiting the level of cellular damage and reducing the
 CC complications of secondary infections.
 XX
 SQ Sequence 1260 BP; 365 A; 236 C; 288 G; 371 T; 0 other;

Query Match 17.0%; Score 65.8; DB 21; Length 1260;
 Best Local Similarity 51.3%; Pred. No. 2.7e-11;
 Matches 179; Conservative 0; Mismatches 167; Indels 3; Gaps 1;
 QY 40 ctcttcgatgctctcttcgcagatctgcgaattgttcttcgaggtcagcgatgctcc 99
 Db 287 ctccctattctcgtgtaatttcgcaattagcgagcaggtaaaagcagtggaagt 346
 QY 100 acttcggcgcgattgggtccacatgcacatcatggtggtggttcttcgccaaattaa 159
 Db 347 tggcagggttgattgattcattgtgtatgctatggcaggtggttcttcgccaaattaa 406
 QY 160 ctattggcgtccagttattgaagttgagaagacacacaaaggcatttggattgttc 219
 Db 407 ctatcgagacctctcg99gttgatgctcttcggcgttgacagatctcttccttggatgttc 466
 QY 220 acctatggttacaaatccctcttgattgttgaaccttggcaagctggtgctctg 279
 Db 467 atctgatgatgggaaccgcagacagagatccggtatttcacaaagcagtgagata 526
 QY 280 gttttacattca---cgtagacacataaaagataactggaagaaacttaccaaagaa 336
 Db 527 ttgttagtgtacattgtgaacagcaatccaccatccatttgcattgcacgtcaatcaaa 586
 QY 337 tcaagtcacatggcatgattcctcgtgtgtagcattaaagcgtggaccgcc 385
 Db 587 taaaagcttagggctaaagctggaggtgttctcaaaccttggaaacccc 635

RESULT 7
 T98644
 ID T98644 standard; DNA; 1234 BP.
 XX
 AC T98644;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE DNA encoding a S. pneumoniae protein of unknown function.
 XX
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis; ss.
 XX
 OS Streptococcus pneumoniae.
 XX

FH Key Location/Qualifiers
 FT CDS 55..933
 XX /*tag= a
 PN WO9743303-A1.
 XX
 XX 20-NOV-1997.
 XX
 PF 14-MAY-1997; 97WO-US07950.
 XX
 PR 14-MAY-1996; 96US-0017670.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI; 1998-008793/01.
 DR P-PSDB; W38594.
 XX
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 XX
 PS Claim 4; Pages 167-168; 483pp; English.
 XX
 CC This sequence encodes a Streptococcus pneumoniae protein of unknown
 CC function, and represents a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 XX
 SQ Sequence 1234 BP; 355 A; 260 C; 278 G; 341 T; 0 other;

Query Match 16.9%; Score 65.4; DB 19; Length 1234;
 Best Local Similarity 55.5%; Pred. No. 3.6e-11;
 Matches 126; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
 QY 33 aaaatagctcttcgatgctctcttcgacttcgcaatttgccttcgaggtcagcgc 92
 Db 956 aagattgctcgtcaattctctggcagcagattatgccaaacttgaactgaaatcaacgt 1015
 QY 93 atgtccacttcggcgcgattggtccacatggacatcatggatggcattttgtcccc 152
 Db 1016 ctagaagcaactggggcagaatagcccatatcgatcatcgacagtcatttgcacg 1075
 QY 153 aatttaactattgctcctccagttattgaaagtttgaaaagcagacaaaggcattttg 212
 Db 1076 caaatcagttttgtgctcaggtgtggtcagagccttcgtctcatagtaagatggtttc 1135
 QY 213 gattgtcaccttattggtttacaaatccctctgtatttatgttaaacctt 259
 Db 1136 gattgccaacttgatgggtgtcaaaccttgaacatcatctctggaaaaatt 1182
 RESULT 8
 T96346/c
 ID T96346 standard; DNA; 1542 BP.

XX Z96346;
 XX 10-APR-2000 (first entry)
 XX S. pneumoniae derived DNA from ORF #174.
 XX Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
 XX Streptococcus pneumoniae.
 XX WO9806734-A1.
 XX 19-FEB-1998.
 XX 15-AUG-1997; 97WO-US14436.
 XX 16-AUG-1996; 96US-0024022.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Stodola RK;
 XX WPI: 1998-159452/14.
 DR P-PSDB; Y86006.
 XX Streptococcus pneumoniae proteins and related DNA - useful for
 PT screening compounds for antibacterial activity
 XX Claim 4; Page 207-208; 640pp; English.
 XX This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see 236173-236494) and their encoded proteins (see
 CC Y85792-Y86182). The DNA, vectors and host cells described in the method
 CC of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.
 XX Sequence 1542 BP; 421 A; 367 C; 328 G; 426 T; 0 other;

Query Match 16.9%; Score 65.4; DB 19; Length 1542;
 Best Local Similarity 55.5%; Pred. No. 4e-11;
 Matches 126; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
 QY 33 aaatatgctccttcgatgctctctccgacttcgcgaatttggttcogaggtcagcgc 92
 DB 279 AAGATTGCTCCGTCATTCGGCACCAGATATGCGCACTTGAACGNGAAATCAACGT 220
 QY 93 atgtccacttcggcgccgattggtccacatggacatggatggatggatggatggatggat 152
 DB 219 CTAGAAGCAACTGGGGCAGAAATATGCCATATCGATATCATGGACAGTCATTTGTACCG 160
 QY 153 aatttaactattggcgctccagttattaaagttagaaggaagacacacacacacacacac 212
 DB 159 CAATCAGTTTTGGTGGAGGTGGTGGAGAGCCCTCGTCCTCATAGTAAGATGTTTC 100
 QY 213 gattgtcaccttatggttacaaatcctcttgattattgtgaacctt 259
 DB 99 GATTGCCACTTGTGTCACAAACCTGAGCATCATCTCGGAAGATT 53

RESULT 9

A26928/C
 ID A26928 standard; DNA; 1019 BP.
 XX A26928;
 AC A26928;
 XX 29-JUN-2000 (first entry)
 XX Essential Staphylococcus aureus gene #79.
 XX Essential gene; Staphylococcus aureus infection; screening; prevention;
 KW antibacterial agent; bacterial infection; ss.
 XX Staphylococcus aureus.
 XX OS
 XX US6037123-A.
 XX 14-MAR-2000.
 XX 13-SEP-1996; 96US-0714918.
 XX 15-SEP-1995; 95US-0003798.
 XX 22-DEC-1995; 95US-0009102.
 XX (MICR-) MICROCID PHARM INC.
 XX Benton B, Lee VJ, Martin PK, Schmid MB, Sun D, Malouin F;
 WPI: 2000-282222/24.
 XX Novel methods for screening for antibacterial agents, useful for
 PT treatment or prevention of Staphylococcus infection, by testing
 PT compounds against, essential bacterial genes -
 XX Disclosure; Column 183-184; 256pp; English.
 XX The present sequence represents an essential Staphylococcus aureus gene.
 CC The invention relates to a method for screening for an antibacterial
 CC agent. The method comprises determining if a test compound is active
 CC against the essential bacterial genes (A26850-A26956). The invention also
 CC includes a method for evaluating an agent for activity on the essential
 CC genes. Also included in the invention is the production of an
 CC antibacterial agent. The antibacterial agents of the invention are used
 CC to treat or prevent bacterial infections, particularly where caused by
 CC Staphylococcus aureus. The antibacterial agents are unlikely to be
 CC affected by known resistance mechanisms.
 XX Sequence 1019 BP; 314 A; 211 C; 151 G; 340 T; 3 other;

Query Match 14.9%; Score 58; DB 21; Length 1019;
 Best Local Similarity 56.9%; Pred. No. 7.6e-09;
 Matches 161; Conservative 0; Mismatches 118; Indels 4; Gaps 3;
 QY 105 ggcgcgattgggtccacatggacatcgatggatggatggatggatggatggatggatggat 164
 DB 1001 GGNITTCACGAGTTCATTTTGGATTTGATGGATGGTCAATTTGCCCTAATATATCTATT 942
 QY 165 ggcgcctcagttattgaagttagaagacacacacacacacacacacacacacacacacac 224
 DB 941 GGTTTACCAATATTAGATGAGTAAGAAAGGACACACATTAACNTATAGACGTACATTG 882
 QY 225 atggtt-acaaatcctcttgattatgttgaccccttggcaaaagctgggtgcttct-gggt 282
 DB 881 ATGATTGAAAATCCAGAAAAGTATATTGTCATCATTTGTCAGAACATGGTGGCGATATGAT 822
 QY 283 ttacatttcacgtagacatcaaaagataaactcggaaagaacttccaaagaatacaagt 342
 DB 821 TTNCATTTATGTGGAATCAACGCCCTCATA--TTTCATCGTCTATTCAATGATTAAAC 764
 QY 343 cacatggcatgattcctcctggttagcattaaagcctggagcccc 385
 DB 763 ATTAGATAAAAAAGCTGGGTGATTAATTAATCTCGGTACACC 721

Db 389 agttggcaggtgattgacacgttgatgaatggtgacggtttgttgcctcccaata 448
 Qy 157 taactattgg 166
 Db 449 ttacaatcgg 458

RESULT 12

V52182
 ID V52182 standard; DNA; 11443 BP.
 AC V52182;
 XX
 XX 23-OCT-1998 (first entry)
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:49.
 XX
 XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 XX WO9818931-A2.
 PN
 XX
 XX 07-MAY-1998.
 PD
 XX
 XX 30-OCT-1997; 97WO-US19588.
 PF
 XX
 XX 31-OCT-1996; 96US-0029960.
 PR
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 PI
 XX
 XX WPI; 1998-272225/24.
 DR
 XX
 XX Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 XX
 XX Claim 1; Page 445-451; 1409pp; English.
 PS
 XX
 XX The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
 CC on it, or a representative fragment or a sequence at least 95% identical
 CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
 CC to 391, identifying members of the library which contain sequences
 CC that hybridize to the target sequence and isolating the nucleic acid
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
 CC from an organism, amplifying nucleic acid molecules whose nucleotide
 CC sequence is homologous to amplification primers derived from the
 CC fragment of the S. pneumoniae genome to prime the amplification and
 CC isolating the amplified sequences. The computer readable medium can be
 CC used in a computer-based system for identifying fragments of the
 CC S. pneumoniae genome of commercial importance, or expression modulating
 CC fragments of the S. pneumoniae genome. Products from the present
 CC invention can be used in diagnosis kits and assays, and pharmaceutical
 CC compositions and vaccines for S. pneumoniae.
 XX
 XX Sequence 11443 BP; 3350 A; 2057 C; 2533 G; 3502 T; 1 other;

Query Match 10.8%; Score 41.8; DB 19; Length 11443;
 Best Local Similarity 47.2%; Pred. No. 0.0037;

Matches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
 Qy 120 cacatggacatcatgagtgccattttgtccccaatttaactattggcgtccagattatt 179
 Db 7739 catatcgattatgagtgccattttgtccccaatttaactattggcgtccagattatt 7798
 Qy 180 gaaagtttgagaaagcacacaaagcatattttgattgtcaccttatgtttacaaatcct 239
 Db 7799 caagaagttcaaaaaattagtgacacacctttatcagttcatctgatggtcacagacca 7858
 Qy 240 ctgtattatgttgaaoccttggcaaaagctggtgttctgtttacattttacacgtagag 299
 Db 7859 accttttgggtagatcaagttctctgatatttacaatgtgagtattttgtattcatctgtaa 7918
 Qy 300 acatcaaaagataactggaaagaacttatccaaagaatcaagtcacatggcattctct 359
 Db 7919 gttctgaatggtctgtctgttttctgttgatgataaaattcatgatgcagggtctcaaggct 7978
 Qy 360 ggtgtagcattaaaagcctgggacccctgt 388
 Db 7979 ggtgtgtcttaactcctgaaacacctgt 8007

RESULT 13

Q27406/C
 ID Q27406 standard; DNA; 6866 BP.
 AC Q27406;
 XX
 XX 05-FEB-1993 (first entry)
 DT
 DE Human parainfluenza virus 2 large protein gene.
 DE
 XX PIV-2; vaccine; ss.
 KW
 XX Human parainfluenza virus-2.
 OS
 XX JP04197179-A.
 PN
 XX
 XX 16-JUL-1992.
 PD
 XX
 XX 29-NOV-1990; 90JP-0325169.
 PF
 XX
 XX 29-NOV-1990; 90JP-0325169.
 PR
 XX
 XX (FUJG) FUJIKURA KASEI KK.
 PA
 XX
 XX WPI; 1992-288437/35.
 DR
 XX
 XX Human para-influenza type 2 virus large protein gene - useful in
 PT diagnosis and as a vaccine against para-influenza virus
 PT
 XX Claim 1; Page 1-4; 17pp; Japanese.

XX Vero cells were cultured in the presence of actinomycin D and
 CC infected with PIV-2. Poly(A)+ RNA was isolated from the infected
 CC cells and a cDNA library was prepared. The library was screened with
 CC a nucleocapsid RNA probe. A positive clone containing the PIV-2
 CC gene was sequenced. The gene can be used for early diagnosis of
 CC PIV-2 infection and as a raw material for vaccines.

XX Sequence 6866 BP; 2257 A; 1274 C; 1194 G; 2141 T; 0 other;

Query Match 9.8%; Score 38.2; DB 13; Length 6866;
 Best Local Similarity 55.7%; Pred. No. 0.04; Mismatches 0; Gaps 0;
 Matches 73; Conservative 0

Qy 209 ttgtgattgtcaccttatggtttacaaatcctcttattgtttgaacccttgcaaaagc 268
 Db 1971 TTTTGGCTTTCCCTTATTTTAGAATCAGCTCTTGATTGTGGAACCTTGGATGAGA 1912
 Qy 269 tgggtctctgtttttacatttcacgtagacatcaaaagaactggaagaacttat 328


```

Db 1911 TGYTATCTCGGATATCTCTTCGCTTTTCTGATATTATGCAATTTGACTCATAGTAA 1852
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 329 ccaagaataca 339
      ||| ||| |||
Db 1851 GCAATGATTTA 1841
      ||| ||| |||

RESULT 14
A31286
ID A31286 standard; DNA; 439 BP.
AC A31286;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #247.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
PN WO9967421-A1.
XX
PD 29-DEC-1999.
XX
PF 25-JUN-1999; 99WO-NZ00092.
XX
PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukkala IJ, Bloksberg LN, Glenn M;
XX
DR WPI; 2000-116958/10.
XX
PT New plant microsatellite markers and associated flanking species for
PT the detection of polymorphic genetic markers -
XX
PS Claim 1; Page 148; 392pp; English.
XX
CC Sequences A31040-A32093 represent novel plant microsatellite sequences
CC and associated flanking species. The sequences comprise a central core
CC repeat sequence, especially selected from the sequences A32094-A32096
CC with left and right flanking sequences. The polynucleotide sequences
CC can be used in the detection of DNA polymorphisms, in genome mapping,
CC in physical mapping, in positional cloning of genes, in variety
CC identification and in evaluation of genetic variability within and
CC between plant tissues, populations, cultivars, species and species
CC groups. They may also be used to design hybridization probes for
CC oligonucleotide fingerprinting and library screening and to design
CC primers for microsatellite-primed PCR. Microsatellite markers are
CC useful to locate specific economically useful genes in plant genomes.
XX
SQ Sequence 439 BP; 122 A; 95 C; 119 G; 102 T; 1 other;

Query Match 9.7%; Score 37.6; DB 21; Length 439;
Best Local Similarity 60.9%; Pred. No. 0.016;
Matches 78; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 40 ctccctcagctctctccgaacttcgc-caatttgcttcgcaggctcagcgatctc 98
      |||| |||| |||| ||| |||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 ctccatcgatctctcagcaaatctcgtactagagaaaccaggtaaaagcagtggat 371
      |||| |||| |||| ||| |||| ||| ||| ||| ||| ||| ||| ||| |||
QY 99 cacttcggcgccgcttgctccacatggacatcatggatggcattttgtcccaattta 158
      |||| |||| |||| ||| |||| |||| |||| |||| |||| |||| ||||
Db 372 gtggcggtgtgactggtatccatgctgatgctgatggcgtctgtttgtccgaatc 431
      |||| |||| |||| ||| |||| |||| |||| |||| |||| |||| ||||
QY 159 actattgg 166

```

```

Db 432 accattgg 439
      || |||||

RESULT 15
X61720
ID X61720 standard; DNA; 573 BP.
XX
AC X61720;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein coding sequence, t457.nt.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12718.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMMUNE INC.
XX
XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX WPI; 1999-189980/16.
XX P-PSDB; Y20023.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
XX Claim 1; Page 165; 275pp; English.
XX
CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
SQ Sequence 573 BP; 208 A; 43 C; 97 G; 225 T; 0 other;

Query Match 9.4%; Score 36.4; DB 20; Length 573;
Best Local Similarity 51.2%; Pred. No. 0.045;
Matches 85; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 172 cagttattgaaagtttgagaaagcacacaaagcattttggattgtcaccttatggtta 231
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 caaaaattcgtgggtgaaatctcaggcaagagatgatttcattttatcctttttt 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 232 caaatcctcttgatttggtgaaccttggtgcaaaagctggtgcttctgttttacattc 291
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 ttaataatctaagatagagataataggtagaaaaaatatttctaagggtttgaatttg 259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 292 acctagagacatacaagataaactggaagaacttatccaaagaat 337
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 aagttgtttaaaaaattataactttcaaaacggtatagaaaaatt 305
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: November 4, 2000, 13:45:16

Sat Nov 4 18:11:21 2000

us-09-300-482-298.rng

Page 10

Job time: 16443 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:33:43 ; Search time 189.35 Seconds
(without alignments)
309.911 Million cell updates/sec

Title: US-09-300-482-298

Perfect score: 388

Sequence: 1 ggagaaagaaagaaagatg.....ttaagcctgggaccccccgt 388

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/6_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	58	14.9	1019	5	US-08-714-918-79
C 2	34	8.8	3781	5	US-08-688-988-5
C 3	31.6	8.1	40352	5	US-08-846-111B-15
C 4	31	8.0	1430	1	US-08-276-452A-25
C 5	31	8.0	1430	3	US-08-798-744-25
C 6	30.6	7.9	3150	5	US-08-673-814-5
C 7	30	7.7	2168	5	US-08-749-522-6
C 8	29.8	7.7	2622	2	US-08-766-014-23
C 9	29.8	7.7	5300	2	US-08-766-014-1
C 10	29.4	7.6	6960	4	US-08-841-349-3
C 11	29.4	7.6	8176	4	US-08-841-349-5
C 12	29	7.5	4182	1	US-07-973-257-1
C 13	29	7.5	5181	2	US-08-257-073-10
C 14	29	7.5	8700	3	US-08-392-625-16
C 15	29	7.5	8700	3	US-08-466-961A-16
C 16	28.4	7.3	1347	2	US-08-663-713A-1
C 17	28.4	7.3	1347	5	US-09-014-888-1
C 18	28.4	7.3	2835	2	US-08-750-532-2
C 19	28.4	7.3	4765	2	US-08-750-532-8
C 20	28.4	7.3	10207	2	US-08-920-812-2
C 21	28.4	7.3	10207	2	US-08-920-812-2
C 22	28.4	7.3	10207	2	US-08-921-177-2
C 23	28.4	7.3	10207	2	US-08-362-577C-2
C 24	28.4	7.3	10207	3	US-08-920-828-2
C 25	28.2	7.3	2465	5	US-08-619-812-5
C 26	28	7.2	2403	2	US-08-441-139-17
C 27	28	7.2	2509	4	US-09-014-969-1
C 28	27.8	7.2	5589	1	US-08-465-795-1
C 29	27.8	7.2	5775	2	US-08-306-691B-15
C 30	27.6	7.2	5775	6	PCT-US93-06251-29
C 31	27.6	7.1	1685	1	US-08-105-483-370
C 32	27.6	7.1	1685	2	US-08-709-209-370
C 33	27.6	7.1	1685	2	US-08-458-101-370
C 34	27.6	7.1	8700	4	US-08-645-193B-18
C 35	27.4	7.1	846	4	US-08-743-637B-189
C 36	27.4	7.1	1631	5	US-09-118-319-1
C 37	27.2	7.0	707	6	PCT-US94-06430-25
C 38	27.2	7.0	1359	6	PCT-US94-06430-21
C 39	27.2	7.0	3828	6	PCT-US93-10500-1
C 40	27.2	7.0	5393	4	US-08-591-079-9
C 41	27.2	7.0	7470	5	US-08-417-089-5
C 42	27.2	7.0	7721	5	US-08-772-270A-14
C 43	26.8	6.9	1759	2	US-08-726-525-5
C 44	26.8	6.9	1759	2	US-08-487-942-5
C 45	26.8	6.9	1759	4	US-08-726-036A-5

ALIGNMENTS

RESULT 1

US-08-714-918-79/c

; Sequence 79, Application US/08714918

; Patent No. 6037123

; GENERAL INFORMATION:

; APPLICANT: Benton, Bret

; APPLICANT: Lee, Ving

; APPLICANT: Malouin, Francois

; APPLICANT: Martin, Patrick K.

; APPLICANT: Schmid, Molly B.

; APPLICANT: Sun, Dongxu

; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL

; NUMBER OF INVENTION: TARGET GENES

; NUMBER OF SEQUENCES: 111

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: 633 West Fifth Street

; CITY: Suite 4700

; STATE: Los Angeles

; COUNTRY: California

; ZIP: U.S.A.

; ZIP: 90071-2066

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

; MEDIUM TYPE: storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: IBM P.C. DOS 5.0

; SOFTWARE: Word Perfect 5.1

; CURRENT APPLICATION DATA:

; FILING DATE: September 13, 1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/009,102

; FILING DATE: December 22, 1995

; APPLICATION NUMBER: 60/003,798

; FILING DATE: September 15, 1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Warburg, Richard J.

; REGISTRATION NUMBER: 32,327

; REFERENCE/DOCKET NUMBER: 222/005

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (213) 489-1600

; TELEFAX: (213) 955-0440

; TELEX: 67-3510

; INFORMATION FOR SEQ ID NO: 79:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1019 base pairs

; TYPE: nucleic acid

Sequence 1, Appli
Sequence 1, Appli
Sequence 15, Appli
Sequence 29, Appli
Sequence 370, App
Sequence 370, App
Sequence 370, App
Sequence 18, Appli
Sequence 189, App
Sequence 1, Appli
Sequence 25, Appli
Sequence 21, Appli
Sequence 9, Appli
Sequence 5, Appli
Sequence 14, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli

; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-714-918-79

 Query Match 14.9%; Score 58; DB 5; Length 1019;
 Best Local Similarity 56.9%; Pred. No. 3.8e-10;
 Matches 161; Conservative 0; Mismatches 118; Indels 4; Gaps 3;

 QY 105 ggcgcgattggtccacatggacatcattggatggcatttggcccaataaactatt 164
 Db 1001 GGNITGACGAGGATTCATTGATGATGATGATGATGATGATGATGATGATGAT 942
 QY 165 ggcgcctcagttattgaaattgaaagacacacaaagggcatatttggattgtcacctt 224
 Db 941 GGTATTACCAATATTAGATGAGTAAAGAAAGGACACACATTACATGATGATGATGAT 882
 QY 225 atggtt-acaaatctcttattgattgtgaacccctggcaaaagctgggtctct-ggtt 282
 Db 881 ATGATTGAAATCCAGAAAGATATATTGCATCATTTGCAGAACATGTTGCCGATATGAT 822
 QY 283 ttacatttcacgtagacatcaaaagataaactggaaagaaacttatccaaagaatcaagt 342
 Db 821 TTNCAATTCAATGTCGAATCAACGCGCTCATA--TTTCATCGCTATTCAAATGATTAAAC 764
 QY 343 cacatggcatgatctcctggttagcattaaagcctgggaacccc 385
 Db 763 ATTTAGATAAAAAGCTGGTGTAGTAATTAATCTCTGTACACC 721

 RESULT 2
 US-08-988-988-5/c
 ; Sequence 5, Application US/08688988B
 ; Patent No. 6096545
 ; GENERAL INFORMATION:
 ; APPLICANT: Lefebvre, Daniel D.
 ; APPLICANT: Malboobi, Mohammad A.
 ; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
 ; FILE REFERENCE: PPL96-03
 ; CURRENT APPLICATION NUMBER: US/08/688.988B
 ; CURRENT FILING DATE: 1996-07-31
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 5
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis Thaliana
 ; US-08-688-988-5

 Query Match 8.1%; Score 31.6; DB 5; Length 40352;
 Best Local Similarity 54.2%; Pred. No. 2.9;
 Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

 QY 206 atattggattgcacattggtttacaaatccctctgatttattgtaacccttggcaaa 265
 Db 13954 AGATGTGGAGGTAGAGACATGATATGATGATGATGATGATGATGATGATGATGAT 14013
 QY 266 agctggtgctctggttttaccatttcacgtagacatcaaaagataaactgggaagaa 323
 Db 14014 TGTGGGTGCTTTTGTCTTCATTTTACAAAAACGTTTAAAAATTTAAAAAGTAAAAA 14071

 RESULT 4
 US-08-276-452A-25/c
 ; Sequence 25, Application US/08276452A
 ; Patent No. 5646029
 ; GENERAL INFORMATION:
 ; APPLICANT: Chen, Chao-Guang
 ; APPLICANT: Mau, Shiao-Lim
 ; APPLICANT: Du, He
 ; APPLICANT: Gane, Allison M
 ; APPLICANT: Bacic, Antony
 ; APPLICANT: Clarke, Adrienne E
 ; TITLE OF INVENTION: Plant Arabinogalactan Protein (AGP) Genes
 ; NUMBER OF SEQUENCES: 91
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee and Winner, P.C.
 ; STREET: 5370 Manhattan Circle, Suite 201
 ; CITY: Boulder
 ; STATE: Colorado
 ; COUNTRY: United States of America
 ; ZIP: 80303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible

; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-714-918-79

 Query Match 14.9%; Score 58; DB 5; Length 1019;
 Best Local Similarity 56.9%; Pred. No. 3.8e-10;
 Matches 161; Conservative 0; Mismatches 118; Indels 4; Gaps 3;

 QY 105 ggcgcgattggtccacatggacatcattggatggcatttggcccaataaactatt 164
 Db 1001 GGNITGACGAGGATTCATTGATGATGATGATGATGATGATGATGATGATGAT 942
 QY 165 ggcgcctcagttattgaaattgaaagacacacaaagggcatatttggattgtcacctt 224
 Db 941 GGTATTACCAATATTAGATGAGTAAAGAAAGGACACACATTACATGATGATGATGAT 882
 QY 225 atggtt-acaaatctcttattgattgtgaacccctggcaaaagctgggtctct-ggtt 282
 Db 881 ATGATTGAAATCCAGAAAGATATATTGCATCATTTGCAGAACATGTTGCCGATATGAT 822
 QY 283 ttacatttcacgtagacatcaaaagataaactggaaagaaacttatccaaagaatcaagt 342
 Db 821 TTNCAATTCAATGTCGAATCAACGCGCTCATA--TTTCATCGCTATTCAAATGATTAAAC 764
 QY 343 cacatggcatgatctcctggttagcattaaagcctgggaacccc 385
 Db 763 ATTTAGATAAAAAGCTGGTGTAGTAATTAATCTCTGTACACC 721

 RESULT 2
 US-08-988-988-5/c
 ; Sequence 5, Application US/08688988B
 ; Patent No. 6096545
 ; GENERAL INFORMATION:
 ; APPLICANT: Lefebvre, Daniel D.
 ; APPLICANT: Malboobi, Mohammad A.
 ; TITLE OF INVENTION: PHOSPHATE STARVATION-INDUCIBLE PROTEINS
 ; FILE REFERENCE: PPL96-03
 ; CURRENT APPLICATION NUMBER: US/08/688.988B
 ; CURRENT FILING DATE: 1996-07-31
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 5
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis Thaliana
 ; US-08-688-988-5

 Query Match 8.8%; Score 34; DB 5; Length 3781;
 Best Local Similarity 57.5%; Pred. No. 0.11;
 Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

 QY 267 gctggtgctctggttttaccatttcacgtagacatcaaaagataaactgggaagaactt 326
 Db 3441 GCTGGTGGCTCTGTTTAACTTAACCTAGTACCTTTTCAAAATAAATAATAGAACTG 3382
 QY 327 atccaaagaatcaagtacatggcattgctcctggttagacattaa 372
 Db 3381 TTGCGAAATGAATGCAATGCTTATTATTTCATCGCAACGTTTAA 3336

 RESULT 3
 US-08-846-111D-15
 ; Sequence 15, Application US/08846111D
 ; Patent No. 6017705
 ; GENERAL INFORMATION:
 ; APPLICANT: Lurquin, Christophe; Brasseur, Francis;
 ; APPLICANT: Boon-Falleur, Thierry
 ; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Are
 ; TITLE OF INVENTION: Members Of The Mage-B Family and Uses Thereof
 ; NUMBER OF SEQUENCES: 18

```

> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: PatentIn Release #1.0, Version #1.25
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/276,452A
> FILING DATE: 18-JUL-1994
> CLASSIFICATION: 435
> ATTORNEY/AGENT INFORMATION:
> NAME: Caruthers, Jennie M.
> REGISTRATION NUMBER: 34,464
> REFERENCE/DOCKET NUMBER: 27-91A
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (303)499-8080
> TELEFAX: (303)499-8089
> TELEX: 49617824
> INFORMATION FOR SEQ ID NO: 25:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1430 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> MOLECULE TYPE: cDNA
> FEATURE:
> NAME/KEY: CDS
> LOCATION: 2..1312
> FEATURE:
> NAME/KEY: misc.feature
> LOCATION: 41..112
> OTHER INFORMATION: /note= "Derived amino acid sequence
> OTHER INFORMATION: corresponding to the peptide sequence by protein
> OTHER INFORMATION: microsequencing"
> FEATURE:
> NAME/KEY: misc.feature
> LOCATION: 41..112
> OTHER INFORMATION: /note= "Derived amino acids 14-24,
> OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
> OTHER INFORMATION: by protein microsequencing"
> FEATURE:
> NAME/KEY: misc.feature
> LOCATION: 25..31
> OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
> OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
> OTHER INFORMATION: 26 can be T instead of A"
US-08-276-452A-25

Query Match      8.0%; Score 31; DB 1; Length 1430;
Best Local Similarity 46.2%; Pred. No. 0.67;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps

QY   143 tttgtcccaatttaactattggcgctccaagttattgaaagttagagaagaacacacaaa 202
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   992 TGTAAGTTCGAGATGAGCATGTGGTTTGTTGTTGAATTCACAGAAAGCATTGT 933
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   203 ggcataattggatgtcaccttattgaacaatcctcttgattgatgtgaaccttcgc 262
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   932 TGTATTGTTTTGTTGTAATTCGCAGAAACAATTATTTGTTGTTAGTTCGCGAGA 873
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   263 aaagctgctgctctggttttacatttcacgtagagacatcaaaagataaactggaaaga 322
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   872 AAGCAATTGTTGTTATTGTTGTTGTAATTCGCGAGAAAACATTTATTGTTATTATTGTTGT 813
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY   323 acttatcaaaagaatcaagtcacatggcatgattccctggcta 365
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   812 TGTAAATTCFCGAGAAAAACATTTATTGTTATTATTGTTGTTGTA 770
     ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

> APPLICANT: Du, He
> APPLICANT: Gane, Alison M
> APPLICANT: Basic, Antony
> APPLICANT: Clarke, Adrienne E
> TITLE OF INVENTION: Plant Arabinogalactan Protein (AGP) Genes
> NUMBER OF SEQUENCES: 91
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: Greenlee and Winner, P.C.
> STREET: 5370 Mannheim Circle, Suite 201
> CITY: Boulder
> STATE: Colorado
> COUNTRY: United States of America
> ZIP: 80303
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Floppy disk
> COMPUTER: IBM PC compatible
> OPERATING SYSTEM: PC-DOS/MS-DOS
> SOFTWARE: Patent in Release #1.0, Version #1.25
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/798,744
> FILING DATE: 13-FEB-1997
> CLASSIFICATION: 435
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 08/276,452
> FILING DATE: 18-JUL-1994
> ATTORNEY/AGENT INFORMATION:
> NAME: Caruthers, Jennie M.
> REGISTRATION NUMBER: 34,464
> REFERENCE/DOCKET NUMBER: 27-91A
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (303)499-8080
> TELEFAX: (303)499-8089
> TELEX: 49617824
> INFORMATION FOR SEQ ID NO: 25:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 1430 base pairs
> TYPE: nucleic acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> MOLECULE TYPE: cDNA
> FEATURE:
> NAME/KEY: CDS
> LOCATION: 2..1312
> FEATURE:
> NAME/KEY: misc-feature
> LOCATION: 41..112
> OTHER INFORMATION: /note= "Derived amino acid sequence
> OTHER INFORMATION: corresponding to the peptide sequence by protein
> OTHER INFORMATION: microsequencing"
> FEATURE:
> NAME/KEY: misc-feature
> LOCATION: 41..112
> OTHER INFORMATION: /note= "Derived amino acids 14-24,
> OTHER INFORMATION: 28, 30, 32-37 are identical to sequences obtained
> OTHER INFORMATION: by protein microsequencing"
> FEATURE:
> NAME/KEY: misc-feature
> LOCATION: 25..31
> OTHER INFORMATION: /note= "Amino acids 25, 27, 29, and
> OTHER INFORMATION: 31 are hydroxylated proline residues; amino acid
> OTHER INFORMATION: 26 can be T instead of A"
> US-08-798-744-25

```

Db 932 TGTATTGTTTGTGTAATTCGGAGAAACAATTATTGTTGTTAGTCTCGAGA 873
Qy 263 aaaagctggctctgttttacatttcagtagacatcaaaagataactggaaaga 322
Db 872 AAGCATGTTGTTATTGTTGTTGTAATTCGGAGAAACAATTATTGTTATTGTTGT 813
Qy 323 acttatcaagaatcaagtcacatgcatgattcctcgtggtga 365
Db 812 TGTAAATTCGGAGAAACAATTATTGTTATTGTTGTTGTA 770

RESULT 6

US-08-673-814-5/c
; Sequence 5, Application US/08673814
; Patent No. 6086894
; GENERAL INFORMATION:
; APPLICANT: Inzana, Thomas J.
; APPLICANT: Ward, Christine
; TITLE OF INVENTION: RECOMBINANT VACCINE FOR DISEASES CAUSED
; TITLE OF INVENTION: BY ENCAPSULATED ORGANISMS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Whitham, Curtis, Whitham & McGinn
; STREET: 11800 Sunrise Valley Dr., Suite 900
; CITY: Reston
; STATE: VA
; COUNTRY: USA
; ZIP: 20191

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/673,814.
; FILING DATE: 27-JUN-1996
; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:
; NAME: Whitham, Michael E.
; REFERENCE/DOCKET NUMBER: VTIP 95-067
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-391-2510
; TELEFAX: 703-391-9035
; TELEX: 283072

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3150 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:

; NAME/KEY: CDS
; LOCATION: 376..1557
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1586..2740
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2743..3150

US-08-673-814-5

Query Match 7.9%; Score 30.6; DB 5; Length 3150;
Best Local Similarity 51.1%; Pred. No. 1.4;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 127 acatcatggatggcgatttgcctcccaatttaactattggcgctccagttattgaaagt 186
Db 1678 AATAACGGATTAGCTTTTAAACCAACTTTTCTGTTTAAACCGGTTTGTAGTTT 1619
Qy 187 tgaagaagacacaaagcgcatattgtgattgtcaccattggttacaaatccctcgtatt 246

Db 1618 TTCTGCGACTTGGCGTAATAGTGTGTTTCCATAATTTTATCCTTATAAATCAATATATT 1559
Qy 247 atgttgaaccccttgqcaaaag 267
Db 1558 AATAGCGAGAAATAGCATTAG 1538

RESULT 7

US-08-749-522-6/c
; Sequence 6, Application US/08749522
; Patent No. 6096950
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal
; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53202-4497

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,522
; FILING DATE:
; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 670513.90244
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2168 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

US-08-749-522-6

Query Match 7.7%; Score 30; DB 5; Length 2168;
Best Local Similarity 53.4%; Pred. No. 1.9;
Matches 63; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 215 ttgtcaccttatggttacaaatccctctgatttatgttgaaaccttggaagctggtgc 274
Db 1247 TTAAACATCTTAATTACATCTTAAATTTATTCATGATTTATATATCAATCAGGTCT 1188
Qy 275 ttctggttttacatttcacgtagacacatcaaaagataactggaaagacttatccaa 332
Db 1187 TTTTGTACTAAATTTGTTATATATACCGTTAAAAAATGACATGTCAGATCTTATGFAA 1130

RESULT 8

US-08-766-014-23
; Sequence 23, Application US/08766014
; Patent No. 574312
; GENERAL INFORMATION:
; APPLICANT: Mamone, Joseph A.
; APPLICANT: Davis, Maria
; APPLICANT: Sha, Dan
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon

Db 3946 ATGGTGGTCATGAAGTCTCTCTGCTTTTAAATGCG 3912

RESULT 11

US-08-841-349-5/c
; Sequence 5, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO4470USO
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 8176
; TYPE: DNA
; ORGANISM: Mus musculus
US-08-841-349-5

Query Match 7.6%; Score 29.4; DB 4; Length 8176;

Best Local Similarity 56.8%; Pred. No. 6.4; Mismatches 0; Gaps 0;
Matches 54; Conservative 0; Indels 41; Indels 0; Gaps 0;

Qy 73 tggcttcgagctcagcgatgctccacttcgcccgcgattggtccacatggacatca 132

Db 4006 TCGCTCACCAGTCTTCGCCAGTCTCCACACACGATTGATCTCTCCTGGTGGCATCC 3947

Qy 133 tggatggcatttggcccccaatttaactattggc 167

Db 3946 ATGGTGGTCATGAAGTCTCTCTGCTTTTAAATGCG 3912

RESULT 12

US-07-973-257-1/c

; Sequence 1, Application US/07973257
; Patent No. 5378820
; GENERAL INFORMATION:
; APPLICANT: Keeler, Jr. Calvin L.
; APPLICANT: Dohms, John E.
; TITLE OF INVENTION: Gene Encoding Cytadhesin
; TITLE OF INVENTION: Protein of Mycoplasma Gallisepticum and Its Use
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly and Hutz
; STREET: 1220 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch diskette

COMPUTER: IBM/PC or Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/973,257

FILING DATE: 19921109

CLASSIFICATION: 435

PRIOR APPLICATION DATA: No. 5378820e

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 4182 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double stranded

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

HYPOTHETICAL: No

ANTI-SENSE: No

ORIGINAL SOURCE:

ORGANISM: Mycoplasma Gallisepticum

; STRAIN: S6

US-07-973-257-1

Query Match

Best Local Similarity 7.5%; Score 29; DB 1; Length 4182;

Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 206 atatttggttgcacattggttacaaatcctctcttattatgtgaaccttggcaaa 265

Db 4102 ATCTTAATATATACCTTAATATTCAATTTCTTAGTGTGCTGGAGCTTTGGTTT 4043

Qy 266 agctggtctctggttttacatttcacgtaga 298

Db 4042 AGCTGGTGTGTTGGTTTTCCTCCTGGAGCAGA 4010

RESULT 13

US-08-257-073-10

; Sequence 10, Application US/08257073

; Patent No. 5766597

; GENERAL INFORMATION:

; APPLICANT: Paoletti, Enzo

; APPLICANT: de Taisne, Charles

; APPLICANT: Tine, John A.

; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.

; STREET: 530 Fifth Avenue, 25th Floor

; CITY: New York

; STATE: New York

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/257,073

; FILING DATE: 09-JUN-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/075,783

; FILING DATE: 11-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/852,305

; FILING DATE: 18-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/672,183

; FILING DATE: 20-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Frommer, William S.

; REGISTRATION NUMBER: 25,506

; REFERENCE/DOCKET NUMBER: 454310-2570

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 840-3333

; TELEFAX: (212) 840-0712

; TELEX: 425066 CURTMS

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5181 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-257-073-10

Query Match

Best Local Similarity 7.5%; Score 29; DB 2; Length 5181;

Matches 62; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Search completed: November 4, 2000, 13:33:58
Job time: 16674 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:10 ; Search time 4352.3 Seconds
(without alignments)
551.186 Million cell updates/sec

Title: US-09-300-482-298

Perfect score: 388

Sequence: 1 ggagagaagaagaaagatg.....ttaagcctggagaccccgct 388

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estba.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estom.*
68: em_estov1.*
69: em_estov2.*
70: em_estpl1.*
71: em_estpl2.*
72: em_estpl3.*
73: em_estpl4.*
74: em_estpl5.*
75: em_estro1.*
76: em_estro2.*
77: em_estro3.*
78: em_estro4.*
79: em_estro5.*
80: em_estro6.*
81: em_estro7.*
82: em_estro8.*
83: em_estro9.*
84: em_estro10.*
85: em_estro11.*
86: em_estro12.*
87: em_estro13.*
88: gb_gss1.*
89: gb_gss2.*
90: gb_gss3.*
91: gb_gss4.*
92: em_gss1.*
93: em_gss2.*
94: em_gss3.*
95: em_gss4.*
96: gb_gss5.*
97: gb_gss6.*
98: gb_gss7.*
99: gb_gss8.*
100: gb_gss9.*
101: em_gss5.*
102: em_gss6.*
103: em_gss7.*
104: em_gss8.*
105: em_gss9.*
106: em_gss10.*
107: em_gss11.*
108: gb_gss10.*
109: gb_gss11.*
110: em_gss12.*
111: gb_gss12.*
112: gb_gss13.*
113: gb_gss14.*
114: gb_gss15.*
115: gb_gss16.*
116: gb_gss17.*

LOCUS
DEFINITION
Sa34C12.y1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl004-1199 5' similar to TR:014105 014105 RIBULOSE-PHOSPHATE
3-EPIMERASE. ; , mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI437973 418 bp mRNA EST 18-APR-2000
AI437973.1 GI:4289768
EST
soybean.
Glycine max

REFERENCE
AUTHORS

1 (bases 1 to 418)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.

TITLE
JOURNAL
COMMENT

Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

FEATURES

1. 418
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl004-1199"
/clone_lib="Gm-cl004"
/tissue_type="root"
/lab_host="Xl10-Gold"
/note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
XhoI; Root cDNA. The mRNA was isolated from entire roots
of 8 day old 'Williams' seedlings which were propagated on
paper towels with distilled water. Stratagene's cDNA
Synthesis Kit (catalog #200401) was used to synthesize the
cDNA. First-strand synthesis was performed with 5-methyl
dCTP, hence the ligated cDNA is hemimethylated.
Stratagene's first-strand synthesis primer was used
[GAGAGAGAGAGAGAGAGAGAGAGAG(T)-18]. After
second-strand synthesis, the cDNA ends were 'polished'
with clone Pfu DNA polymerase, ligated to EcoRI adapters,
and phosphorylated. The XhoI site within the first-strand
synthesis primer was restricted by digestion with XhoI;
all XhoI sites in the cDNA would be protected by their
hemimethylated status. The cDNA constructs were
size-fractionated with a 500bp cutoff, using GIBCOBRL Life
Technologies' cDNA Size fractionation column. The column
eluent was then ligated into Stratagene's pBluescript II
XR Predigested vector (pBluescript II SK(+)) that had been
digested with EcoRI and XhoI, and phosphorylated. Both
the white and blue colonies appear to contain recombinant
plasmids with cDNA inserts. Blue colonies 9n-15) have been
sequenced, and possess putative cDNA inserts. This library
was constructed by Dr. Paul Keim & Virginia H. Coryell,
Department of Biology, Box 5640, Northern Arizona
University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr.
Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax:
520-523-7500, email: paul.keim@naui.edu,

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	365.4	94.2	418	10	AI437973 sa34c12.y
2	320.8	82.7	403	18	AV412302
3	310.4	80.0	405	10	AI442690 sa85b04.y
4	307.8	79.3	641	23	AW682881 NF001C10L
5	296.8	76.5	633	34	BE325053 NF119G01S
6	285.4	73.6	736	21	AW348508 GM210002B
7	281.2	72.5	399	18	AV425192
8	281.2	72.5	448	14	AL371120 MLCBA42A06
9	244.6	63.0	825	25	AW983286 HVSMEG001
10	243	62.6	494	35	BE404347 WHE0441.B
11	238.2	61.4	526	36	BE494864 WHE1258.D
12	237.4	61.2	407	15	AO069503
13	226.8	58.5	483	19	AV554048
14	226	58.2	497	11	AI490023 EST248362
15	226	58.2	552	12	AI772355 EST253455
16	226	58.2	593	19	AW031337 EST274791
17	226	58.2	630	20	AW223738 EST300549
18	226	58.2	653	13	AI896020 EST265463
19	225	58.0	483	35	BE353381 EST353662
20	225	58.0	552	11	AI485294 EST243598
21	224.4	57.8	520	12	AI771373 EST252473
22	222.4	57.3	569	19	AV553860 AV553860
23	222	57.2	451	36	D41947 RICS4976A.R
24	221.8	57.2	508	11	AI485741
25	220.6	56.9	450	13	AI898953
26	220	56.7	378	9	AI164978
27	212.4	54.7	434	9	AI164426
28	211.8	54.6	522	9	AI162283
29	207.4	53.5	570	9	AI161491
30	203	52.3	380	35	BE358283
31	201.6	52.0	348	35	BE404517
32	198.4	51.1	457	36	D41569
33	197.6	50.9	352	36	D24955
34	187.8	48.4	725	19	AW010209 ST03004.P
35	174	44.8	272	9	AI164995 A072p43u
36	172.4	44.4	638	21	AW350186 GM210007B
37	158.4	40.8	613	5	AW556420 275 Loblo
38	152	39.2	268	9	AI165092
39	150.2	38.7	441	9	AI161745 A006p24u
40	148.8	38.4	209	9	AI165543 A085p48u
41	147	37.9	544	19	AV545174 AV545174
42	128.4	33.1	232	36	D41516
43	116.8	30.1	749	33	BE034126 MG05C09.M
44	101.8	26.2	170	20	AW224026 EST300837
45	101	26.0	465	36	C88722 C88722 Mous

ALIGNMENTS

```

BASE COUNT      111 a 93 c 101 g 113 t
ORIGIN          virginia.coryell@nau.edu"

```

Query Match	94.2%	Score 365.4;	DB 10;	Length 418;
Best Local Similarity	97.1%;	Pred. No. 3.3e-99;		
Matches 372;	Conservative	0;	Mismatches 11;	Indels 0; Gaps 0;

Qy	6	aagaagaagaatgggaatgacacccgaataatgctctcttcgatgtctcttcgcacttc	65
Db	4	AGNAGGAGGATGGGAGTGACACGGAATAATGCTCTTCGATGTCTCTTCGACTTC	63
Qy	66	gccaatattggcttcgcgaggtcagcgactgtctccacttcgcgccgatggctccacatg	125
Db	64	GCCAAATTGGCTTCGAGSCTCAGGCGATGCTCCACTTCGGGCCGCGATTGGCTCCACATG	123
Qy	126	gacatcatggatggcgacatttggccccaatctaactattgcgctccagttattgaaagt	185
Db	124	GACATCATGGATGGCGATTTTGTGCCAAATTAACATTGGCGCTCCAGTTATTGAAAGT	183
Qy	186	ttgagaaagcacacaaaggcatatttgatgtcaccttatgtgtacaaatcctttgat	245
Db	184	TTGAGAAAGCACACAAAGGATATTTCGATTGTCCACCTTAATGTTTCAAAATCCTCTTGAT	243
Qy	246	tatgttgacccttggcgaagctggctgtctctgtgtttacatttcacgtagagacatca	305
Db	244	TATGTTTGAACCCCTTGCGAAAAGCTGGTGCTCTGTTTACATTTTCATGTAGAGACATCA	303
Qy	306	aaagataaactggaaagaacttatccaagaatacaagtccacgtgatgtattcctgggtata	365
Db	304	AAAGATAACTGGGAAGAACTTATCCAAGAATCAAGTCACATGGCATGACTCCTGGTGTA	363
Qy	366	gcattaaagctgggacccccggt	388
Db	364	GCATTAAAGCCTGGGACCCCAT	386

RESULT	2
AV412302	
LOCUS	AV412302 403 bp mRNA EST 23-MAY-2000
DEFINITION	Japonicus Lotus japonicus young plants (two-week old) Lotus
ACCESSION	japonicus cDNA clone M9217c10_r 5', mRNA sequence.
VERSION	AV412302.1 GI:7741466
KEYWORDS	EST.
SOURCE	Lotus japonicus.
ORGANISM	Lotus japonicus Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus. 1 (bases 1 to 403) Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S. Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus DNA Res. 7 (2), 127-130 (2000) 20277479 Contact: Yasukazu Nakamura The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/ . Location/Qualifiers 1 .403
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
COMMENT	
FEATURES	
source	

```

/organism="Lotus japonicus"
/db_xref="taxon:34305"
/clone="MMW21/cu0_t"
/clone_lib="Lotus japonicus young plants (two-week old)"
/dev_stage="young plants (two-week old)"
/notes="vector: pBluescriptII SK- ; Site_1: EcoRI; Site_2:
xhoI; isolate=Miyakojima MG-20"
113 a 93 c 91 g 106 t

```

Query Match	82.7%;	Score 320.8;	DB 18;	Length 403;	
Best Local Similarity	91.4%;	Pred. No. 7.9e-86;			
Matches 340;	Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0;
QY	4	gaagaagaagaatgggaatgacacgcagcaaaaatagctctctcgatgctctctcttcgcgact	63		
Db	32	GTAAACAACAACGATGGGAGGTGACACCCAAAATTGCTCCGTCGATGCTCATCGGACT	91		
QY	64	tcgccaattggcttcgcgaggctcgcagctgctccacttcggcgccgatggctccaca	123		
Db	92	TCGCCAATTGGCTTCCGAGGCTCAACGCATGCTTGATTATGGCGCCGATTGGCTCCACA	151		
QY	124	tggacatcatggatggcgatttggcccccaatttaactatbaggcgtccagtattgaaa	183		
Db	152	TGGACATCATGATGGGCACCTTTGCCCCAAATTAACTATTGGCACTCCGGCTATTGAAA	211		
QY	184	gttgagaaagcacaaaggcattttggattgtcacctatggttacaatactcttg	243		
Db	212	GTTTGAGAAAGCACAAAGGCATATCTGGATTGTCACCTTATGGTTACAATTCCTCTTG	271		
QY	244	attatgtgaaccttggcaaaagctggcgcttcgtgtttacatttcacgtagagacat	303		
Db	272	ATTATGTTGAGCCCTTGGGAAAGCAGGTGCTTCTGGTTTTACATTTTCATGTAGAGGCAT	331		
QY	304	caaaagataactggaaagaacttatccaaaagaatacagtcacatggcatgttcctagt	363		
Db	332	CAAAAGATAACTGGAAAGAAGCTTATCCNAAGAAATTAAGTCACAGGGCATGAAGCCTGGTG	391		
QY	364	tagcattaaagc	375		
Db	392	TAGCATTAAAGC	403		

RESULT 3
AI442690
LOCUS
DEFINITION
AI442690 405 bp mRNA EST 18-APR-2000
s885804.y1 Cm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Cm-cl004-0036 5' similar to TR:O14105 O14105 RIBULOSE-PHOSPHATE
3-EPIMERASE. ; mRNA sequence.
ACCESSION
VERSION
AI442690
KEYWORDS
AI442690.1 GI:4297942
EST.
SOURCE
glycine max
ORGANISM
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 405)
AUTHORS
Shoemaker, R., Kelm, P., Vodkin, L., Erpelding, J., Corvett, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers,
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert Length: 923 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 395

POLYA-NO.

FEATURES

source

Location/Qualifiers

1. .405

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl004-6056"

/clone_lib="Gm-cl004"

/tissue_type="root"

/lab_host="Xil10-Gold"

/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. StrataGene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. StrataGene's first-strand synthesis primer was used [GAGAGAGAGAGAGAGACTAGTCTCGAG(T)-18]. After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into StrataGene's pBluescript II XR predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n=15 have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@naau.edu, virginia.coryell@naau.edu"

BASE COUNT

111 a 86 c 96 g 112 t

ORIGIN

Query Match 80.0%; Score 310.4; DB 10; Length 405;

Best Local Similarity 91.6%; Pred. No. 1.le-82;

Matches 348; Conservative 0; Mismatches 11; Indels 21; Gaps 1;

QY 9 aaagaaaatgggaatgacaccgaaatagctcttcgatgtctctccgactcgcc 68
|||||
Db 1 AAAGGAAGATGGAGTGACACCAGAAATTGCTCTTCGATGCTCTTCGCATTCGCC 60

QY 69 aatttgcttcaggagctcagcgcatctcccacttcgcgcgcatgttgctccacatggac 128
|||||
Db 61 AATTGGCTTCGAGGCTCAGCGCAT-----GCTCCACATGGAC 99

QY 129 atcatgtagggcattttgtcccccaatttaactatttggtggtccagtattgaaagttag 188
|||||
Db 100 ATCATGTGATGGCGATTTTCTCCCAATTAACATAATTTGGCGCTCCAGTTAATGAAAAGTTTG 159

QY 189 agaaagcacacaagaagcatatttggtgattgaccttatgtttacaaaactctctgattat 248
|||||
Db 160 AGAAAGCACACAAGGGATATTGGATTGTCACCTTATGTTTACAAATCCTCTTGATTAT 219

QY 249 gttagaaccttggcaaaagctgggtcttgttttacatttcacagtagagacatcaaaa 308
|||||
Db 220 GTTAGCGCTTGGCAAAGCTGGTGCTGCTGTTTTACATTTTCATGTAGAGACATCAAAA 279

QY 309 gataaactggaagaacttatccaagaatcaaagcatggtcattgattcttggttagca 368
|||||
Db 280 GATAAATGGGAAGAAGTATTCAAGAATCAAGTCACATGGCATGGCATGCTCTTGTTAGCA 339

QY 369 ttaaagcctgggacccocgt 388
|||||
Db 340 TTAAGCGCTGGGACCCCCAT 359

RESULT 4

AW682881

LOCUS

DEFINITION

NF001C10LF1081 Developing leaf Medicago truncatula cDNA clone NF001C10LF 5', mRNA sequence.

ACCESSION

AW682881

VERSION

AW682881.1

GI:7557597

EST.

barrel medic.

KEYWORDS

Medicago truncatula

SOURCE

ORGANISM

Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weiler,W. and May,G.D.

AUTHORS

Expressed Sequence Tags from the Samuel Roberts Noble Foundation

TITLE

Medicago truncatula leaf library

JOURNAL

Unpublished (2000)

COMMENT

Contact: May GD

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7391

Fax: 580 221 7380

Email: gdmay@srbnf.org

Insert Length: 641

Std Error: 0.00

Plate: 001

Row: C

Column: 10

Seq primer: TCACACAGGAACACGCTATGAC.

LOCATION/QUALIFIERS

1. .641

/organism="Medicago truncatula"

/db_xref="taxon:3880"

/clone="NF001C10LF"

/clone_lib="Developing leaf"

/tissue_type="leaf"

/dev_stage="pooled developmental"

/note="Vector: Lambda Zap; Contains a mixture of very young, developing, mature and senescing leaves."

BASE COUNT

201 a 130 c 139 g 171 t

ORIGIN

Query Match 79.3%; Score 307.8; DB 23; Length 641;

Best Local Similarity 87.7%; Pred. No. 7e-82;

Matches 336; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 6 aagaagaaagatgggaatgacaccgaaatagctcttcgatgtctctccgacttc 65
|||||
Db 126 AAAAGGAAGAAGATGGGATGACACCGGAAANTASCTCTTCGATGTCATCAGATTTT 185

QY 66 gccaatgttgcttcaggagctcagcgcatgctccacttcgggcgcatggctccacatcg 125
|||||
Db 186 GCTAATTTGGCTTCGGAAGCTCATCGTATGATCAATTAATGCGGCTGATTGGCTTCATG 245

QY 126 gatcatgtagtggtggaattttgtcccaattttaactattggcgtccagttattgaaagt 185
|||||
Db 246 GATATCATGATGGGCATTTTGTCCCTTAATTAACATATGGCGCTCCAATCATTTGAAAGT 305

QY 186 ttgagaagcacacaaaggcatatttggtgattgtcaccttatggtttacaaatcctcttgat 245
|||||
Db 306 TTGAGAAGCACACAGAGGCATATCTGGACTGTCACTGTATGGTTACAAATCCACTTGAT 365

QY 246 tatgtgaaccttggcaaaagctgggtctgtctgttttattcatttcacgttagacatca 305
|||||
Db 366 TATGTTGAACCTTTGGGAAAAGCTGGTGCCTCTGGTTTTTACATTTTATATAGACATCA 425

QY 306 aaagataactggaaagaacttatccaagaatcaaagcatggtcattgattcttggtgta 365
|||||
Db 426 AAAGACAACCTGGAAGAAGTATATCCAAAAATTAAGTCACACGCGCATGAGGCGCTGGTGA 485

[illegible]

QY	64	tgcccaatttgctccgagctcagcgcatgctccacttcggcgccgatggctccaca	123
Db	134	tgcccaatttggcttcgagcgctacgcgatgcttgattatggcgccgatggctccaca	193
QY	124	tggacatcatggtggcgatttgcgccaaatttaactattggcgctccagttattgaaa	183
Db	194	tggacatcatggtggcgactttgtccccaatttaactatttggcactcgggtcattgaaa	253
QY	184	gtttgagaagcacacaaggcattattggattgtgcacottatggttcacaatcctcttg	243
Db	254	gtttcagaaagcacacaaaggcatatctggattgtcaccttatgttacaattcctcttgc	313
QY	244	attatgttgaaaccttggcaaaagctgctcttgattttacatttcacgttagacat	303
Db	314	attatgttgagcccttgggaaaagcaggtgcttctggttttacctttcatgttagagggcat	373
QY	304	caaaagataacttggaagaaccttacc	329
Db	374	caaaagataacttggaagaaccttacc	399

RESULT 8

AL371120

LOCUS

DEFINITION

448 bp

mtBa Medicago truncatula

EST

cdna clone

03-AUG-2000

MLBA42A06 T3, mRNA

sequence.

KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
AUTHORS	1 (bases 1 to 448) Journet,E.P., Crespeau,H., van-Tulnen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianluazzi-Pearson,V. and Gamas,P.
TITLE	Medicago truncatula ESTs from nitrogen-starved roots
JOURNAL	Unpublished (2000)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: secret@genoscope.cns.fr , Web : www.genoscope.cns.fr Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES	Location/Qualifiers 1, 448 /organism="Medicago truncatula" /cultivar="Jemalong" /db_xref="taxon:3880"

```

/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MLBA42A06"
/clone_lib="MtBA"
/tissue_type="root tips"
/dev_stage="harvested after 3 days of N-starvation"
/notes="vector: plusscript psk; Site_1: EcoRI; Site_2:
XhoI; plants were grown in an aeroponic chamber for 14
days on nitrogen-rich medium followed by 3 days on N-free
medium. RNA was extracted from root tips (1-3 cm). cDNA
was prepared from polyA+ enriched RNA. The cDNA was

```

extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Séquençage (Genoscope, Evry, France)."

TITLE
JOURNAL
COMMENT

/organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE0441_B06_C11"
 /clone_lib="Wheat etiolated seedling root cDNA library"
 /tissue_type="Root"
 /dev_stage="Five day old etiolated seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site.1: EcoRI; Site.2: XhoI; Seeds were surface-sterilized
 , germinated and grown aseptically in the dark at room
 temperature on filter paper with water, nystatin and
 cerotaxime in covered crystallization dishes. Roots were
 harvested. The tissue, total RNA, and poly(A) RNA were
 prepared, a cDNA library was made, and the cDNA clones
 were in vivo excised to give pBluescript phagemids in the
 TJ Close lab (Choi, Close, Fenton) at the University of
 California, Riverside. Plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."
 113 a 127 c 143 g 111 t

BASE COUNT 113 a 127 c 143 g 111 t
ORIGIN

Query Match 62.6%; Score 243; DB 35; Length 494;
 Best Local Similarity 79.3%; Pred. No. 1.9e-62;
 Matches 288; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 26 gacacgaaatagctcttcgactgtcttcgacttcgcaatttggcttcgagggc 85
 DB 53 GCGCGGAGATAGCGCGTCCATGCTGCTCGGACTTCGCCAACCTCGCTCGGAGGC 112
 QY 86 tcagcgatgtccacttcggcgccgattggtccacatgagacatcatgatggcattt 145
 DB 113 CGAGCGCATGCTCGGCTCGGCGCGACTGGCTCCAGATGACGTCATGATGGGCACT 172
 QY 146 tgcceaatatgaattgctccagttattgaaattgagaaagcacacaaagggc 205
 DB 173 TGTTCCTAATTACTATGGAGCTCCAGTATGAGAGCTTGAGAGGACACACAAAGGC 232
 QY 206 atatttgattgacaccttattgataaaatcctcttgcatttattgaaaccttggcaaa 265
 DB 233 ATATTGAGCTGACATCATGTTGACAAATCCTTCAGATTATGAGGCAATTTGGAAA 292
 QY 266 agctggtctctgttttaacatttcaogtagagacatcaaaagataactggaagact 325
 DB 293 AGCTGGTCCCTCAGGATTCACGATTCATATAGAAAGTAACAGGGATTAACGGAAAGAGCT 352
 QY 326 tatcaaaagaatcaagtcacatggtcattctctggttagcattaaagcctgggacccc 385
 DB 353 CATCCAAAGCATTAAGCAAGGGCATCGACCTGGTGTATCATTAAGCCTGGTACTCC 412
 QY 386 cgt 388
 DB 413 TGT 415

RESULT 11
 BE494864 526 bp mRNA EST 02-AUG-2000
 LOCUS WHE1258_D10_H20S Secale cereale anther cDNA library Secale cereale
 DEFINITION cDNA clone WHE1258_D10_H20, mRNA sequence.
 ACCESSION BE494864
 VERSION BE494864.1 GI:9661373
 KEYWORDS EST.
 SOURCE rye.
 ORGANISM Secale cereale
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Secale.
 REFERENCE 1 (bases 1 to 526)
 AUTHORS Anderson, O.D., Butler, E., Chao, S., Choi, D.W., Close, T.J., Fenton
 R.D., Gustafson, J.P., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R.,
 Miller, R., Rausch, C.J., Ross, K., Seaton, C.I. and Tong, J.C.

The structure and function of the expressed portion of the wheat
 genomes - Anther cDNA library from rye
 Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Stragatene SK primer.

FEATURES

source

Location/Qualifiers

1..526

/organism="Secale cereale"

/cultivar="Blanco"

/db_xref="taxon:4550"

/clone="WHE1258_D10_H20"

/clone_lib="Secale cereale anther cDNA library"

/tissue_type="Anther"

/dev_stage="Adult plant before anthesis"

/lab_host="E. coli SOLR"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site.1: EcoRI; Site.2: XhoI; Plants were grown in the
 greenhouse. Anthers were harvested and pooled from early
 meiosis to late meiosis. The tissue, total RNA, and
 poly(A) RNA were prepared (Butler, Ross and Gustafson) at
 University of Missouri, Columbia. A cDNA library was
 made, and the cDNA clones were in vivo excised to give
 pBluescript phagemids in the TJ Close lab (Choi, Close,
 Fenton) at the University of California, Riverside.
 Plasmid DNA preparations and DNA sequencing were performed
 in the OD Anderson lab (all other authors)."
 123 a 130 c 157 g 116 t

BASE COUNT 123 a 130 c 157 g 116 t
ORIGINQuery Match 61.4%; Score 238.2; DB 36; Length 526;
Best Local Similarity 78.5%; Pred. No. 5.3e-61;
Matches 285; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 26 gacacgaaatagctcttcgactgtcttcgacttcgcaatttggcttcgagggc 85
 DB 12 GCGCGGAGATAGCGCGTCCATGCTGCTCGGACTTCGCCAACCTCGCTCGGAGGC 71
 QY 86 tcagcgatgtccacttcggcgccgattggtccacatgagacatcatgatggcattt 145
 DB 72 CGAGCGCATGCTCGGCTCGGCGCGACTGGCTCCAGATGACGTCATGAGGCAATTTGGAA 131
 QY 146 tgcceaatatgaattgctccagttattgaaattgagaaagcacacaaagggc 205
 DB 132 CGTTCCTAATTACTATGGAGCTCCAGTATGAGAGCTTGAGAGGCAATTTGGGAA 251
 QY 206 atatttgattgacaccttattgataaaatcctcttgcatttattgaaaccttggcaaa 265
 DB 192 ATATTGAGCTGACATCATGTTGTCACAAATCCTTCTGATTATGATAGGCAATTTGGGAA 251
 QY 266 agctggtctctgttttaacatttcaogtagagacatcaaaagataactggaagact 325
 DB 252 AGCTGGCGCTTCAGGATTCACATTCCTATAGAAAGTAGCAGGAAACACTGGAAAGAGCT 311
 QY 326 tatcaaaagaatcaagtcacatggtcattctctggttagcattaaagcctgggacccc 385
 DB 312 CATCCAAAGCATTAAGCAAGGGCATCGGCTGGTGTATCATTAAGCCTGGTACTCC 371
 QY 386 cgt 388
 DB 372 TGT 374

RESULT 12
AU069503

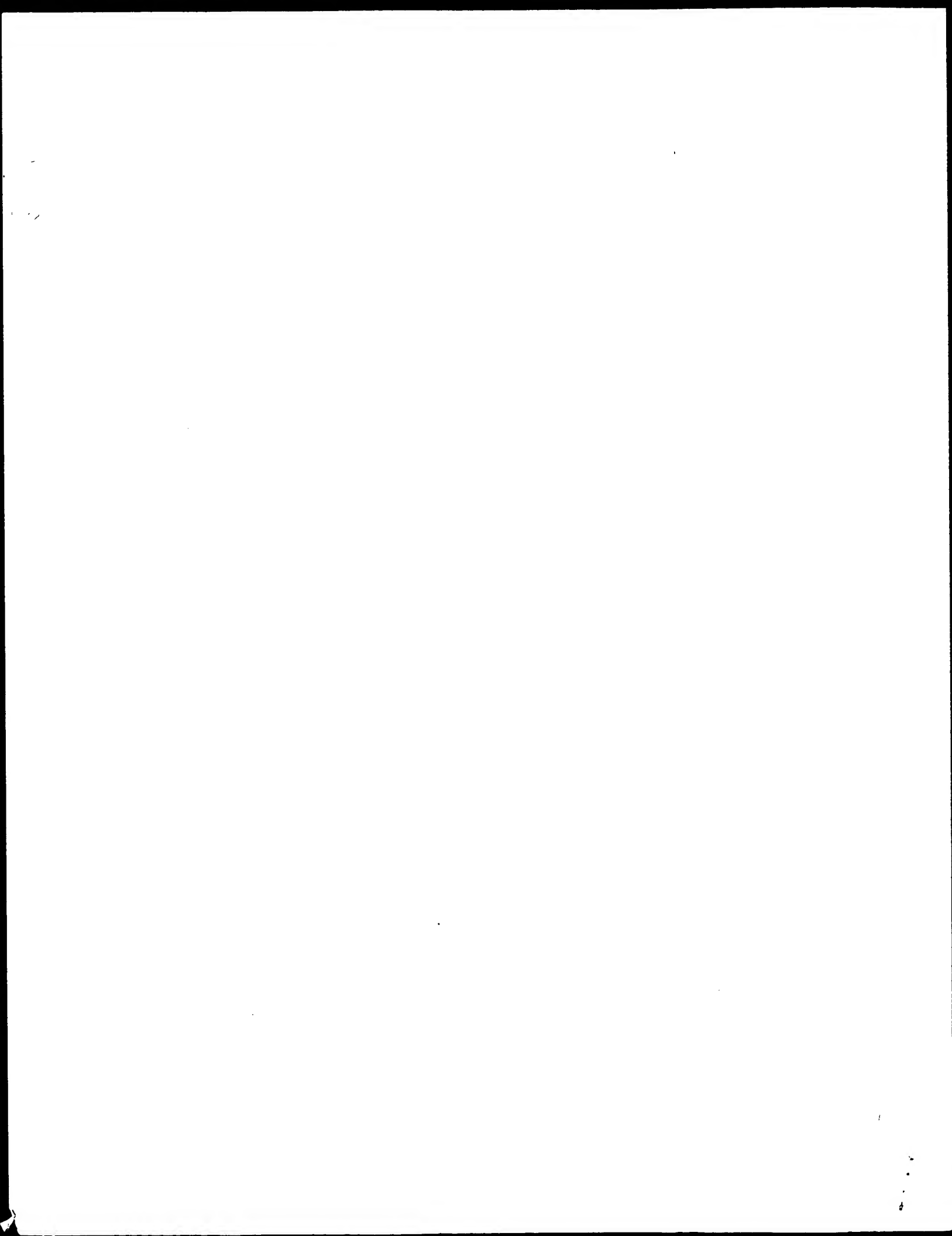
LOCUS AV069503 407 bp mRNA EST 07-JUN-1999
 DEFINITION AV069503 Rice callus Oryza sativa cDNA clone C62255_11A, mRNA sequence.
 ACCESSION AV069503
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
 REFERENCE 1 (bases 1 to 407)
 AUTHORS Yamamoto,K. and Sasaki,T.
 TITLE Rice cDNA from callus (1998)
 JOURNAL Unpublished (1998)
 COMMENT Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai,Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7458
 Email: tsasaki@ab.affrc.go.jp
 PROJECT = "RGP".
 Location/Qualifiers
 1. 407
 /organism="Oryza sativa"
 /strain="cultivar Nipponbare, sub_species Japonica"
 /db_xref="taxon:4530"
 /clone="C62255_11A"
 /clone_lib="Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."
 99 a 106 c 116 g 86 t
 BASE COUNT 99 a 106 c 116 g 86 t
 ORIGIN
 Query Match 61.28; Score 237.4; DB 15; Length 407;
 Best Local Similarity 78.8%; Pred. No. 8.7e-61;
 Matches 283; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
 Qy 26 gacaccgaataagctccttcgactgctcttcgacttcgcaattgcttcccaaggc 85
 Db 49 GCGGCGAAGATACGCGCTCGATGCTCTCGTGGACTTCGCCAACCTCCCGCGGAGGC 108
 Qy 86 tcagcgatgctcacttcgagccgctgattggctccacatgacatcatgagtgccattt 145
 Db 109 CGACCGCATGGTCCGCTCGCGCGCGACTGGCTCCACATGGACATCATGGACGGCATT 168
 Qy 146 tgtcccaatttaactatggcgctccagttattgaagttagaagttgagaagacacacaaggc 205
 Db 169 TGTTCCTTAATCTACTATTGGAGCTCCAGTGTTCAGAGCTTGAGGAAGCACACCAAGGC 228
 Qy 206 atattggattgcacattatggttaacaaatcctcttgattgattgtaaccccttgcaaa 265
 Db 229 ATATTGGACTGCCATCTTATGTGTGACCAATCTTCGGATTATGTAGAACCATTAGCAAA 288
 Qy 266 agctgggtcttctggtttacattcagctagagacatcaaaagataaactggaaagaaact 325
 Db 289 AGCTGGTGGCTCAGGTTTCATTCATTCATATAGAAGTATCCAGAGACAATTTGGCAAGAACT 348
 Qy 326 tatccaagaatacaagtcacatggcattccttggttagcattaaagcctgggacc 384
 Db 349 CATCCAAGTATCAACAAAGGTATGCGACCGGGTGTATCATTTGAGGCCAGGCACCTC 407
 RESULT 13
 AV554048
 LOCUS
 DEFINITION AV554048 Arabidopsis thaliana roots Columbia Arabidopsis thaliana cDNA clone RZ79e04R 5', mRNA sequence.
 EST 23-JUL-2000

ACCESSION AV554048
 VERSION AV554048.1 GI:8725461
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 483)
 AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 Location/Qualifiers
 1. 483
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="RZ79e04R"
 /clone_lib="Arabidopsis thaliana roots Columbia"
 /tissue_type="roots"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
 121 a 108 c 132 g 122 t
 BASE COUNT 121 a 108 c 132 g 122 t
 ORIGIN
 Query Match 58.5%; Score 226.8; DB 19; Length 483;
 Best Local Similarity 75.4%; Pred. No. 1.4e-57;
 Matches 282; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
 Qy 15 agatgggaatgacaccgaaataatagctcttcgactgctcttcgacttcgcaatttg 74
 Db 21 AAGAAGATGGTGTGCGCAAAATAGCACCGTCGATGCTGTCATCGGACTTCGCAATCTA 80
 Qy 75 gcttcgagctcagcgatgctccacttcgagcgcgagtggtgcacatgacatcatg 134
 Db 81 GCGGCGAGGCGCAATCGGATGATCGATTGGCGCGCAACTGGCTTCACATGGATATTG 140
 Qy 135 gatggcatttgcctcccaatttaactattgctcctccagttatgaaagtttgaaaag 194
 Db 141 GACGGSCATTTGTCCCTAATCTTAGGATTGTTGCTCTCTGTCATCGAGATTGCGAAG 200
 Qy 195 cacacaaagccatatttggtgattgcaccttatggttaacaaatcctcttgattatgtgaa 254
 Db 201 CACACTAATAGCATATCTTGTATGCCACCTATGTTGACGAACCCCATGATACCTGCA 260
 Qy 255 ccttggcgaagcctggtcttcgttttattacatttcacgttagacacataaagataac 314
 Db 261 CAGATGGCTTAAAGCTGGGCTTCTGTTTCACATTCCAGCTCGAGTGGCCCCAAGATAAT 320
 Qy 315 tgggaagaacttatccaaagaatacaagtcacatggcattgattcctggttagcattaaag 374
 Db 321 TGGCAACAACTTGTGAGAGAGATTAAAGTCTACTGGGATGAGACAGGCGTGGCTCTAAG 380
 Qy 375 cctgggaccccccgt 388
 Db 381 CCTGGAACACCTGT 394
 RESULT 14
 AV554048
 LOCUS
 DEFINITION AV554048 tomato ovary, TAMU Lycopersicon esculentum cDNA clone CLED19K11, mRNA sequence.
 EST 29-JUN-1999
 ACCESSION AV554048
 VERSION AV554048.1 GI:4385394

KEYWORDS	EST.	AI772355	AI772355	552 bp	mrna	EST	29-JUN-1999
SOURCE	tomato.	LOCUS	EST253455	tomato resistant,	Cornell Lycopersicon esculentum	cdna	
ORGANISM	Lycopersicon esculentum	DEFINITION	clone CLER2M16, mrna sequence.				
REFERENCE	1 (bases 1 to 497)	ACCESSION	AI772355				
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D., and Giovannoni,J.	VERSION	AI772355.1	GI:5270396			
TITLE	Generation of ESTs from tomato carpel tissue	KEYWORDS	EST.				
JOURNAL	Unpublished (1999)	SOURCE	tomato.				
COMMENT	Contact: David Frisch Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU	ORGANISM	Lycopersicon esculentum				
FEATURES	Location/Qualifiers	REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.				
source	1. .497	AUTHORS	1 (bases 1 to 552) D'Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D., Giovannoni,J.J. and Martin,G.B.				
	/organism="Lycopersicon esculentum"	TITLE	Generation of ESTs from Pseudomonas resistant tomato				
	/cultivar="TA496"	JOURNAL	Unpublished (1999)				
	/db_xref="taxon:4081"	COMMENT	Contact: David Frisch Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 4366 Fax: 864 656 4293 Email: dfrisch@CLEMSON.EDU				
	/clone="cLED19K11"	FEATURES	5 prime sequence.				
	/clone_lib="tomato ovary, TAMU"	source	1. .552				
	/tissue_type="carpel"		/organism="Lycopersicon esculentum"				
	/dev_stage="5 days pre-anthesis to 5 days post-anthesis"		/cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"				
	/lab_host="X11-Blue MRF"		/db_xref="taxon:4081"				
	/notes="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: XhoI; cLED - Tomato Carpel EST Library. OligoRT-primed and directionally cloned cDNA in vector Lambda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."		/clone="CLER2M16"				
BASE COUNT	138 a 114 c 125 g 120 t		/clone_lib="tomato resistant, Cornell"				
ORIGIN			/tissue_type="leaf"				
			/dev_stage="4-week old"				
			/lab_host="SOLR"				
			/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: XhoI; CLER - Tomato Pseudomonas Resistant EST Library. Directionally cloned cDNAs inserted into pBlueScript SK(-) at 5' end with EcoRI and 3' end with XhoI site."				
Query Match	58.2%; Score 226; DB 11; Length 497;						
Best Local Similarity	74.9%; Pred. No. 2.4e-57;						
Matches	283; Conservative 0; Mismatches 95; Indels 0; Gaps 0;						
QY	8 gaaagaaagatgggaatgacacccgaaataagctcttcgacttccttcgacttcgc 67						
DB	78 GCAAGTGAGGAGATGGTGAAGGCAATCATAGCACCGTCGATGCTCATCGGACTTGC 137						
QY	68 caatttgcttcgaggtcagcgatgctccacttcggcgccgattggctccacatgga 127						
DB	138 TAATTTGGCTCTGAAGCAGAGCGCATGCTCAATTGTGGTGGGATGGCTCCATATGGA 197						
QY	128 catcatgagggcaattttgtcccaatttaactattggcgctcagttattgaaagttt 187						
DB	198 CATCATGAGCGGTCACTTTGTGCCAAACCTTACCGTCGGTCTCCAAATATCGAGAGTCT 257						
QY	188 gagaagacacaaaggcatatttgatgttcacatttatggttacaaatcctcttgatta 247						
DB	258 GAGAAGCATACAAAGGATATCTGGACTGCCACCTATATGTCACCAACCCCTTGATTA 317						
QY	248 tgttgaaccttggcaaaagcttggctctgtgtttacattacacgttagagacatacaa 307						
DB	318 TGTGGAACCGCTTGGCAAAGCTGGTGCCCTCAGGCTTTACGTTCCATATGAGGATCTAG 377						
QY	308 agataactggaagaaccttaccagaatcaagtcacatgcatgattctctggtgtagc 367						
DB	378 AGATAATTTGGCAAGAGCTTTGTCACAGGATATAAGTCTAAGGCGATGAACCTGGGTTTC 437						
QY	368 attaaagcttgggacccc 385						
DB	438 TTTGAAGCCTGTACACC 455						
RESULT	15						

QY 368 attaaagcctgggaccc 385
|| ||||| || ||
Db 406 TTGAAGCCTGGTACACC 423

Search completed: November 4, 2000, 11:52:13
Job time: 18250 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:44:57 ; Search time 320.8 Seconds
(without alignments)
370.042 Million cell updates/sec

Title: US-09-300-482-225

Perfect score: 316

Sequence: 1 gataaggtgcacactgag.....gaatggaggagtagtacttoga 316

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_36.*
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	46.2	14.6	1246	21 229148	Corn zeaxanthin ep
2	44.2	14.0	1050	21 245413	D-ribulose-5-phosp
3	35.8	11.3	1260	21 245412	D-ribulose-5-phosp
4	34.4	10.9	533	16 T20487	Human gene signatu
C 5	34.4	10.9	613	21 280305	Human colon cancer
C 6	34.4	10.9	667	21 280304	Human colon cancer
C 7	34.4	10.8	2930	17 T42174	Maize id gene. Ze
C 8	34.4	10.8	3669	19 V41721	Maize id gene. Ze
C 9	34.4	10.8	3693	20 Z34479	Maize id gene. Ze
C 10	33.8	10.7	2639	17 T06481	Cystathionine gamm
C 11	33.4	10.6	5511	20 Z08718	Chimeric gene cons
12	32.8	10.4	1149	20 Z21189	Zea mays pathogene

13	32	10.1	500	19 V63934	Mycobacterium tube
14	32	10.1	500	20 X81041	Nucleotide sequenc
C 15	32	10.1	2050	19 V63935	Mycobacterium tube
C 16	32	10.1	2050	20 X81042	Nucleotide sequenc
C 17	32	10.1	16885	17 T33535	BCG deletion regio
18	31.8	10.1	5511	20 Z08718	Chimeric gene cons
C 19	31.4	9.9	9848	20 X20503	Polynucleotide seq
C 20	31	9.8	341	21 Z43056	Human 5' EST isola
C 21	30.2	9.6	300	21 A00751	Human colon cancer
C 22	30.2	9.6	6138	21 Z57864	Protein regulating
C 23	29.8	9.4	622	18 X30872	Streptococcus pneu
C 24	28.2	9.2	1149	20 Z21189	Zea mays pathogene
C 25	28.8	9.1	8585	11 Q50749	Microspore-specifi
C 26	28.8	9.1	8585	18 T48867	Brassica napus mic
C 27	28.8	9.1	17082	20 X13166	Enterococcus faeca
C 28	28.4	9.0	116277	20 X20249	Borrelia burgdorfe
C 29	28.2	8.9	3475	13 Q21453	Sequence encoding
C 30	28.2	8.9	10329	21 Z50264	Genomic DNA of pot
C 31	28.2	8.9	10329	21 Z50653	Genomic DNA encodi
C 32	28.2	8.9	12566	19 V52282	Streptococcus pneu
C 33	28	8.9	8793	21 Z89350	Murine embryo macr
C 34	27.8	8.8	232	20 V90238	EST clone DF971.
C 35	27.8	8.8	249	8 N70901	ORF 8 from MYMV co
C 36	27.8	8.8	2723	8 N70893	MYMV strand (b).
C 37	27.8	8.8	3305	21 A26925	Essential Staphylo
C 38	27.8	8.8	4360	18 T44068	Rat von Ebner's gl
C 39	27.8	8.8	5836	16 Q96296	Human prostate-sp
C 40	27.8	8.8	5836	18 T84444	Human prostate-spe
C 41	27.8	8.8	5836	21 Z99932	DNA sequence of co
C 42	27.8	8.8	7646	18 V74432	Staphylococcus aur
C 43	27.6	8.7	937	13 Q29266	Human calcium chan
C 44	27.6	8.7	1803	13 Q26409	Toxoplasma gondii
C 45	27.4	8.7	756	20 X98830	Human validated ca

ALIGNMENTS

RESULT 1
Z29148/c
ID Z29148 standard; cDNA; 1246 BP.
XX AC Z29148;
XX AC Z29148;
DT 21-FEB-2000 (first entry)
XX DE Corn zeaxanthin epoxidase C-terminal half encoding cDNA.
XX DE Corn zeaxanthin epoxidase; carotenoid biosynthetic enzyme;
KW chimeric gene; transformed cell; inhibitor; antibody; ds.
KW Zea mays.
OS Zea mays.
XX FH Location/Qualifiers
XX FT 2..949
XX FT /*tag= a
XX FT /product= "zeaxanthin epoxidase C-terminal half"
XX PN WO9955889-A2.
XX PD 04-NOV-1999.
XX PF 22-APR-1999; 99WO-US08789.
XX PR 24-APR-1998; 98US-0083042.
XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX PI Cahoon RE, Hitz WD, Shen JB, Williams ME;
DR WPI: 2000-072184/06.
DR P-PSDB; Y44220.


```

DR WPI; 1996-497621/49.
XX P-PSDB; W03698.
PT New isolated plant Id gene - used to develop prods. for use in
PT altering the induction of flowering in plants
XX Claim 6; Fig 2 and Fig 3; 58pp; English.
XX The Id gene controls flower evocation in maize plants. The
CC maize nucleic acid is similar to that of genes encoding zinc-
CC finger regulatory proteins in animals.
CC Transposons Ac and Ds constitute a family of related transposable
CC elements present in maize. A derivative of Ds, Ds2, can be used
CC to produce a new mutant of the Id gene. The Ds2 (in the presence
CC of active Ac) is excised from a nearby gene on chromosome 1
CC and inserted into the Id gene to produce Id*.
XX Sequence 2930 BP; 850 A; 664 C; 643 G; 771 T; 2 other;
SQ

Query Match 10.8%; Score 34; DB 17; Length 2930;
Best Local Similarity 74.1%; Pred. No. 0.19;
Matches 43; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 235 ccggttttttttattcgctgcttttagttcaaacatgaactagcgacgactgat 292
Db 710 CCTTCTTTTATCAGTGGCGCATTTAGTTTAAAAATCACTAGCGGACGATAAATAT 653

RESULT 8
ID V41721/c
XX V41721 standard; DNA; 3669 BP.
AC V41721;
XX
XX 16-NOV-1998 (first entry)
DE Maize Id gene.
XX ss; maize; Id; floral induction; transgenic plant.
XX Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 12..2961
FT exon /*tag= a
FT /*product= "Id protein"
FT 12..228
FT /*tag= b
FT /number= 1
FT 229..318
FT /*tag= c
FT /number= 1
FT exon 319..609
FT /*tag= d
FT /number= 2
FT 380..442
FT /*tag= e
FT /note= "Zinc finger motif"
FT intron 610..728
FT /*tag= f
FT 729..902
FT exon /*tag= g
FT /number= 3
FT 796..858
FT misc_feature /*tag= h
FT /note= "Zinc finger motif"
FT intron 903..2327
FT /*tag= i
FT /number= 3
FT exon 2328..2959
FT /*tag= j

```

```

FT /number= 4
FT 3179
FT /*tag= k
PN W09837201-A1.
XX 27-AUG-1998.
XX 18-FEB-1998; 98WO-US03161.
XX 30-DEC-1997; 97US-0000640.
XX 20-FEB-1997; 97US-0804104.
XX (COLD-) COLD SPRING HARBOR LAB.
XX Colasanti JJ, Sundaresan V;
XX WPI; 1998-467564/40.
XX P-PSDB; W59836.
XX Id gene controlling floral induction in maize - useful to create
PT transgenic plants with earlier, delayed or inhibited floral
PT induction e.g. to extend geographical range of crops
XX Claim 1; Fig 2; 68pp; English.
XX The maize Id gene controls floral induction and is useful in the
CC production of transgenic plants with altered floral induction times,
CC i.e. earlier, delayed or inhibited floral induction. The Id gene and
CC zinc-finger regulatory regions are useful to identify these genes in
CC maize and isolate similar genes in other plants. The polypeptides
CC and antibodies are also useful in Id detection e.g. to locate activity in
CC plants.
XX Sequence 3669 BP; 1011 A; 922 C; 851 G; 884 T; 1 other;
SQ

Query Match 10.8%; Score 34; DB 19; Length 3669;
Best Local Similarity 74.1%; Pred. No. 0.21;
Matches 43; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 235 ccggttttttttattcgctgcttttagttcaaacatgaactagcgacgactgat 292
Db 1437 CCTTCTTTTATCAGTGGCGCATTTAGTTTAAAAATCACTAGCGGACGATAAATAT 1380

RESULT 9
ID Z34479/c
XX Z34479 standard; DNA; 3693 BP.
AC Z34479;
XX
XX 01-FEB-2000 (first entry)
DE Maize Id gene.
XX Indeterminate gene; Id gene; maize; corn; flower induction;
XX floral evocation; transgenic plant; ss.
XX Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 12..2957
FT /*tag= a
FT /note= "contains introns"
FT exon 12..240
FT /*tag= b
FT /number= 1
FT intron 241..329
FT /*tag= c
FT /number= 1
FT exon 330..627
FT /*tag= d

```



```

FH Key Location/Qualifiers
FT CDS 49..468
FT FT /*tag= a
XX WO9844119-A1.
XX 08-OCT-1998.
XX 01-APR-1998; 98WO-DK00132.
XX 05-JAN-1998; 98US-0070488.
XX 02-APR-1997; 97DK-0000376.
XX 18-APR-1997; 97US-0044624.
XX 10-NOV-1997; 97DK-0001277.
XX (STAT-) STATENS SERUM INST.
XX PA
XX PI Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;
XX PI Rosenkrands I, Weldingh K;
XX XX
XX WPI: 1998-542705/46.
XX P-PSDB; W72910.
XX
XX New isolated mycobacteria polypeptides and nucleic acids - used for
XX developing products for the diagnosis of or vaccination against
XX mycobacterial infections, particularly tuberculosis
XX PT
XX PS Claim 23; Page 178-179; 163pp; English.
XX CC
XX The present sequence encodes a Mycobacterium tuberculosis protein.
XX Products from the present invention, which describes protein fragments
XX and nucleic acid fragments derived from M. tuberculosis, can be used in
XX the detection of and prevention of mycobacterial infections. In
XX particular, the proteins and nucleic acids can be used for the diagnosis
XX of or vaccination against tuberculosis caused by M. tuberculosis,
XX M. africanum or M. bovis.
XX CC
XX Sequence 500 BP; 54 A; 111 C; 235 G; 100 T; 0 other;
XX SQ
Query Match 10.1%; Score 32; DB 19; Length 500;
Best Local Similarity 55.4%; Pred. No. 0.39;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 31 ccttccttgacatagagttgaatggtctaggtctcttcaaccatagacgtgcccga 90
Db 163 ccggcgatggcgatgggggtgtgggtgttttagatgctggcggttccggagggcggc 222
QY 91 tctgctggggccaattgcacgtcgctgctggaagctctatatattggcgctgcgg 142
Db 223 tctctcgcgatcgatcgatcgagttggtggcgagggaggtgggtgggg 274
RESULT 14
X81041
ID X81041 standard; DNA; 500 BP.
XX
XX X81041;
XX
XX 06-SEP-1999 (first entry)
XX
XX Nucleotide sequence of antigen RD1-ORF8.
XX
XX Immunogenic; Mycobacterium tuberculosis; immune response; infection;
XX tuberculosis; fusion polypeptide; T-cell epitope; ESAT-6; MPT59; TB;
XX pharmaceutical; vaccination; M. africanum; M. bovis; CFP7A; CFP30A;
XX CFP7B; CFP19; CFP27; CFP30A; RD1-ORF; CFP10A; CFP16; CFP19; CFP23;
XX CFP25A; CFP30B; CFP7B; ss.
XX
XX Mycobacterium tuberculosis.
XX OS
XX WO9924577-A1.
XX

```

```

PD 20-MAY-1999.
XX
XX 08-OCT-1998; 98WO-DK00438.
XX
XX 01-APR-1998; 98WO-DK00132.
XX 10-NOV-1997; 97DK-0001277.
XX 05-JAN-1998; 98US-0070488.
XX (STAT-) STATENS SERUM INST.
XX PA
XX PI Andersen P, Skjot R;
XX PI WPI: 1999-347282/29.
XX P-PSDB; Y21927.
XX
XX New Immunogenic fragment of Mycobacterium tuberculosis
XX Example 2; Page 197-198; 265pp; English.
XX
XX The invention describes a substantially pure immunogenic polypeptide
XX fragment (I) from Mycobacterium tuberculosis that is able to evoke a
XX protective immune response against infections by mycobacteria belonging
XX to the tuberculosis complex. The invention provides a (1) fusion
XX polypeptide comprising at least one polypeptide fragment (I) and at least
XX one fusion partner; (2) a fusion polypeptide fragment comprising a T-cell
XX epitope from M. tuberculosis protein ESAT-6, or MPT59 and a second
XX different amino acid sequence from M. tuberculosis, and/or including a
XX sequence which protects the first amino acid sequence from in vivo
XX degradation or post-translational processing; (3) a nucleic acid fragment
XX that encodes the above polypeptides. The polypeptides and nucleic acid
XX are useful as pharmaceuticals, for diagnosis of and as antigens for
XX vaccination against TB caused by Mycobacterium tuberculosis, africanum or
XX bovis. The polypeptides are also useful for diagnosing ongoing or
XX previous sensitization in an animal with bacteria belonging to the
XX tuberculosis complex. The invention also describes the use of CFP7A or
XX CFP30A or a T-cell epitope of for the induction of a strong immune
XX response in a mammal; use of CFP7B, CFP19 or MPT59-ESAT6 or a T-cell
XX epitope of for diagnosis of TB in a mammal by performing a DTH type skin
XX test; use of CFP27, CFP30A, RD1-ORF2, RD1-ORF3, RD1-ORF5, MPT59-ESAT6,
XX ESAT6-MPT59, CFP10A, CFP16, CFP19, CFP23, CFP25A, CFP30B, CFP7B or a T-
XX cell epitope of for the preparation of an immunological composition; and
XX for the preparation of a subunit vaccine.
XX
XX Sequence 500 BP; 54 A; 111 C; 235 G; 100 T; 0 other;
XX SQ
Query Match 10.1%; Score 32; DB 20; Length 500;
Best Local Similarity 55.4%; Pred. No. 0.39;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 31 ccttccttgacatagagttggtggtggtctaggtcttcaaccatagacgtgcccga 90
Db 163 ccggcgatggcgatgggggtgtgggtgttttagatgctggcggttccggagggcggc 222
QY 91 tctgctggggccaattgcacgtcgctgctggaagctctatatattggcgctgcgg 142
Db 223 tctctcgcgatcgatcgatcgagttggtggcgagggaggtgggtgggg 274
RESULT 15
V63935/c
ID V63935 standard; DNA; 2050 BP.
XX
XX V63935;
XX
XX 21-JAN-1999 (first entry)
XX
XX Mycobacterium tuberculosis antigen RD1-ORF9B encoding DNA.
XX
XX Mycobacterium tuberculosis; antigen; vaccine; immunological;
XX immunogen; infection; ds.
XX
XX Mycobacterium tuberculosis.
XX OS

```

```

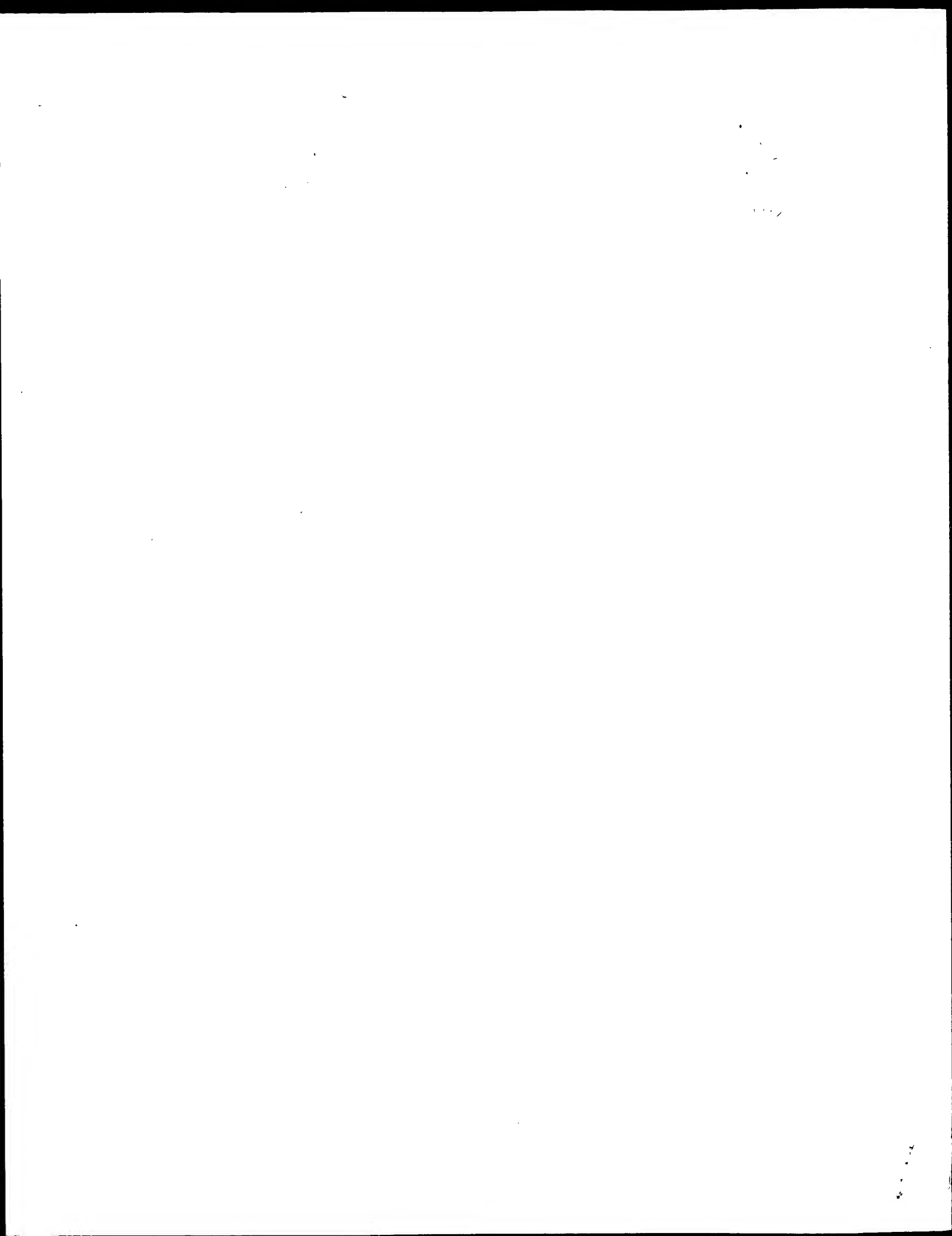
XX Key Location/Qualifiers
FH CDS 22..2022
FT /*tag= a
XX
XX WO9844119-Al.
XX
XX 08-OCT-1998.
XX
XX 01-APR-1998; 98WO-DK00132.
XX
XX 05-JAN-1998; 98US-0070488.
XX
XX 02-APR-1997; 97DK-0000376.
XX
XX 18-APR-1997; 97US-0044624.
XX
XX 10-NOV-1997; 97DK-0001277.
XX
XX (STAT-) STATENS SERUM INST.
XX
XX Andersen P, Florio W, Nielsen R, Oettinger T, Rasmussen PB;
XX Rosenkrands I, Weldingh K;
XX
XX WPI: 1998-542705/46.
XX P-PSDB; W72911.
XX
XX New isolated mycobacteria polypeptides and nucleic acids - used for
XX developing products for the diagnosis of or vaccination against
XX mycobacterial infections, particularly tuberculosis
XX
XX Claim 23; Page 180-183; 163pp; English.
XX
XX The present sequence encodes a Mycobacterium tuberculosis protein.
XX Products from the present invention, which describes protein fragments
XX and nucleic acid fragments derived from M.tuberculosis, can be used in
XX the detection of and prevention of mycobacterial infections. In
XX particular, the proteins and nucleic acids can be used for the diagnosis
XX of or vaccination against tuberculosis caused by M. tuberculosis,
XX M. africanum or M. bovis.
XX
XX Sequence 2050 BP; 404 A; 774 C; 582 G; 290 T; 0 other;

Query Match 10.1%; Score 32; DB 19; Length 2050;
Best Local Similarity 55.4%; Pred. No. 0.74;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 31 ccttcccttagacatagaggttggtgtgtcttagcttcacacatagacgtgcccga 90
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 CCGGCGATGGGATGGGGGTGGGGTTAGATCGGCCCGGTTCGGCGAGGGCGGC 279
QY 91 ttgtctggggccaattgcatcgtcgtggaagctctatatattgctgctgagg 142
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 TCTCTCGCGGATCGGATCGGAGTTGGCGGAGCGGAGGTGGGGTGGGG 227

Search completed: November 4, 2000, 13:45:01
Job time: 16428 sec

```



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: November 4, 2000, 13:33:34 ; Search time 189.35 Seconds
(without alignments)
252.402 Million cell updates/sec

Title: US-09-300-482-225
Sequence: 1 gataaggtgcacactgag.....gaatggaggagacttoga 316

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.8	10.7	3639	4	US-08-737-524B-26
2	32	10.1	16885	2	US-08-390-878-16
3	29.2	9.2	6343	5	US-08-581-148C-30
4	28.8	9.1	8585	1	US-08-030-096-3
5	28.2	8.9	3475	2	US-07-960-389-1
6	27.8	8.8	3305	5	US-08-714-918-76
7	27.8	8.8	4360	1	US-08-470-350B-1
8	27.8	8.8	5836	1	US-08-380-916-1
9	27.8	8.8	5836	5	US-08-721-690-1
10	27.6	8.7	1803	5	US-08-458-922-2
11	27.2	8.6	2375	1	US-08-468-853-9
12	27.2	8.6	2375	2	US-08-468-855-9
13	27.2	8.6	2375	2	US-08-310-357-9
14	27.2	8.6	2375	2	US-08-468-852-9
15	27.2	8.6	2375	4	US-08-468-857-9
16	27	8.5	5035	4	US-08-616-392C-3
17	26.8	8.5	1950	4	US-08-377-440A-2
18	26.8	8.5	3765	5	US-07-705-490-1
19	26.8	8.5	4362	3	US-08-455-073A-1
20	26.8	8.5	28804	3	US-08-592-874-1
21	26.8	8.5	28804	5	US-09-096-942-2
22	26.8	8.5	28804	5	US-09-096-867-2
23	26.8	8.5	29604	5	US-08-781-891-207
24	26.6	8.4	572	2	US-08-253-155A-15
25	26.4	8.4	2631	2	US-08-785-429-1
26	26.4	8.4	2631	5	US-08-996-621-1

c 27	26.4	8.4	40352	5	US-08-846-111D-15	Sequence 15, Appl
c 28	26	8.2	2688	1	US-08-088-633-3	Sequence 3, Appl
c 29	26	8.2	2688	1	US-08-245-756-3	Sequence 3, Appl
c 30	26	8.2	2688	2	US-08-441-750-3	Sequence 3, Appl
c 31	26	8.2	2688	3	US-08-441-751-3	Sequence 3, Appl
c 32	26	8.2	2688	6	PCT-US92-02521-3	Sequence 1, Appl
c 33	25.8	8.2	1478	5	US-08-817-926-1	Sequence 1, Appl
c 34	25.8	8.2	2322	2	US-08-618-164-1	Sequence 1, Appl
c 35	25.8	8.2	3562	5	US-08-817-926-19	Sequence 1, Appl
c 36	25.6	8.1	428	5	US-08-755-587-10	Sequence 19, Appl
c 37	25.6	8.1	501	5	US-08-755-587-9	Sequence 10, Appl
c 38	25.6	8.1	1095	3	US-08-180-524-2	Sequence 9, Appl
c 39	25.6	8.1	1095	4	US-08-975-166-2	Sequence 2, Appl
c 40	25.6	8.1	1129	5	US-08-755-587-30	Sequence 2, Appl
c 41	25.6	8.1	1540	2	US-08-463-115-2	Sequence 30, Appl
c 42	25.6	8.1	1540	2	US-08-465-388-2	Sequence 2, Appl
c 43	25.6	8.1	2381	1	US-08-021-608D-9	Sequence 2, Appl
c 44	25.6	8.1	2381	2	US-08-726-160-9	Sequence 9, Appl
c 45	25.6	8.1	2381	6	PCT-US94-01782-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-737-524B-26
; Sequence 26, Application US/08737524B
; Patent No. 5912414
; GENERAL INFORMATION:
; APPLICANT: CARL SAVERIO FALCO
; APPLICANT: DOMINICK ANTHONY GUIDA, JR.
; APPLICANT: MARY ELIZABETH HARNETT LOCKE
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS, CHIMERIC
; TITLE OF INVENTION: GENES AND METHODS FOR INCREASING
; TITLE OF INVENTION: THE METHIONINE CONTENT OF THE SEEDS
; TITLE OF INVENTION: OF PLANTS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD FOR WINDOWS 95 (7.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08737,524B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: LYNNE M. CHRISTENBURY
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: BB-1059-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3639 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-737-524B-26

Query Match 10.7%; Score 33.8; DB 4; Length 3639;
Best Local Similarity 68.1%; Pred. No. 0.033;


```
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-714-918-76

Query Match      8.8%; Score 27.8; DB 5; Length 3305;
Best Local Similarity 51.2%; Pred. No. 4.5;
Matches 65; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 180 cgagggtctcagaaacaaactgatttgggtttctgtgttaaagtactccctccgtt 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 789 CTAGGTATAAGCAACAATGCGCAACGTGTGATTATAGTCTTCTAAGTATTGACTTTTAT 730

Qy 240 ttttttattctgcgtgtttttagttaaactgaactagcggagcactgatattcgagaa 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 729 CATTGTGTTTCAAAGCGTAACTCTTTGAAAAATTTACTAGCTGACATATAACTTGC AAAAT 670

Qy 300 tggaggg 306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 669 TAAACGG 663

RESULT 7
US-08-470-350B-1/c
; Sequence 1, Application US/08470350B
; Patent No. 5684126
; GENERAL INFORMATION:
; APPLICANT: Li, Xiao
; APPLICANT: Snyder, Solomon H
; TITLE OF INVENTION: Ebnerin: A Secreted von Ebner's Gland
; TITLE OF INVENTION: Protein Associated with Taste Buds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,350B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wolfe, Susan A
; REGISTRATION NUMBER: 33,568
; REFERENCE/DOCKET NUMBER: 01107.48790
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4360 base pairs
; TYPE: nucleic acid

; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-714-918-76

Query Match      8.8%; Score 27.8; DB 1; Length 4360;
Best Local Similarity 55.6%; Pred. No. 5.2;
Matches 74; Conservative 0; Mismatches 57; Indels 2; Gaps 1;

Qy 78 agacgtggccgcacatctgtgtgggccaattgcacatcgctcgctggagctcttatattggcgc 137
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3867 AAAGATGGCTGCAGTGGGATAGACCTCTCTGGAGCTGGCTTCTCCACATCTGCCAC 3808

Qy 138 tgcggaccacagagccatcatatctgtcgtgaggaagagc--gtcggaggctctcagaac 195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3807 TCCCAAGTCGAGACTCCTCTTTTCTTCTGCTGGAGATTGCAACTGGAGGTCCCAAGAC 3748

Qy 196 aaaaactgatttt 208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3747 AACATCCACCTTT 3735

RESULT 8
US-08-380-916-1
; Sequence 1, Application US/08380916
; Patent No. 5648478
; GENERAL INFORMATION:
; APPLICANT: Calydon, Inc.
; TITLE OF INVENTION: Tissue-Specific Enhancer Active in
; TITLE OF INVENTION: Prostrate
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/380,916
; FILING DATE: 12-JAN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/182,247
; FILING DATE: 13-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: FP-60058-PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5836 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MEDIUM TYPE: DNA (genomic)
; US-08-380-916-1
```

```
Query Match      8.8%; Score 27.8; DB 1; Length 5836;
Best Local Similarity 53.2%; Pred. No. 6.1;
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 166 ctgaggaagagcgtcgagggctctcagacaaactgatttgggtgttctgctgtaaa 225
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4839 CTGGGTGACAGAGTGAGACTCTGCTCAAAAAAATAATTTTTTTTTTTTGTAGA 4898

QY 226 gtactccctcgcttttttttttttttttttttttttttttttttttttttttttt 276
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4899 GATGGATCTGTGCTTTCTGCTTGGCTTGAACCTCGGCTTCAAGT 4949

RESULT 9
US-08-721-690-1
; Sequence 1, Application US/08721690
; Patent No. 6057299
; GENERAL INFORMATION:
; APPLICANT: Henderson, Daniel R.
; TITLE OF INVENTION: TISSUE-SPECIFIC ENHANCER ACTIVE
; TITLE OF INVENTION: IN PROSTATE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,690
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/380,916
; FILING DATE: 30-JAN-1995
; APPLICATION NUMBER: US 08/182,247
; FILING DATE: 13-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Catherine, Polizzi M
; REGISTRATION NUMBER: 40,130
; REFERENCE/DOCKET NUMBER: 34802-20001.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5836 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-721-690-1

Query Match      8.8%; Score 27.8; DB 5; Length 5836;
Best Local Similarity 53.2%; Pred. No. 6.1;
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 166 ctgaggaagagcgtcgagggctctcagacaaactgatttgggtgttctgctgtaaa 225
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4839 CTGGGTGACAGAGTGAGACTCTGCTCAAAAAAATAATTTTTTTTTTTTGTAGA 4898

QY 226 gtactccctcgcttttttttttttttttttttttttttttttttttttttttttt 276
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4899 GATGGATCTGTGCTTTCTGCTTGGCTTGAACCTCGGCTTCAAGT 4949
```

```
RESULT 10
US-08-458-922-2/c
; Sequence 2, Application US/08458922
; Patent No. 6077690
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cloning and expression of a protein antigen
; TITLE OF INVENTION: of toxoplasma gondii
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,922
; FILING DATE: 02-June-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: B3125C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-458-922-2

Query Match      8.7%; Score 27.6; DB 5; Length 1803;
Best Local Similarity 49.3%; Pred. No. 3.8;
Matches 72; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 32 ctctcccttgacatagagttgattggtgtcttagtgccttcaaccatagacgtgcccgcac 91
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1736 CTTCTCCGGCTGAGATGTTCCCGATGGTGGCGCCACCCCTCTCATGCCGTTTCCCAT 1677

QY 92 ctgctggggcccaattgcctgcctggaagctcttatatttggcgctgcgacccagag 151
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1676 CAGTTTGATACAGGGGCAAAAGGCTGATAGCTCTGTGCGCAGTTGCTTCTGCTACTCAGAAG 1617

QY 152 ccatcatatctgtgctgaggaagac 177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1616 TCTCCATGGCTTGCAATGGGAGGAGC 1591

RESULT 11
US-08-468-853-9
; Sequence 9, Application US/08468853
; Patent No. 5670362
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5670362el Patent Department
```


CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/310,357
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/904,075
FILING DATE:
APPLICATION NUMBER: EP 91.201.523.7
FILING DATE: 18-Jun-1991
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2375 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
IMMEDIATE SOURCE:
CLONE: Em100E
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1859
US-08-310-357-9

Query Match 8.6%; Score 27.2; DB 2; Length 2375;
Best Local Similarity 52.7%; Pred. No. 6.1;
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 47 aggttgatgggtgcttagctctcaaccatagacggtgcccgcacatctgctggggccaatt 106
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1576 AGATTGAAGCTGTGACAGGACGTACACAGAAAAACAAAGCAAGTATCTGGATCTAAGG 1635
QY 107 gcacgtcgtcgtggaagctctatatattggcgtcgtcgaccaccagagccatcat 158
|||| || || || || || || || || || || || || || || || || || || || || ||
Db 1636 TAGGGTGGCGCTGCTAGTGTGCTAGAGGTGGCTATGCTTATAAAGGAACCTCCTT 1687

RESULT 14
US-08-468-852-9
; Sequence 9, Application US/08468852
; Patent No. 5792644
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5792644el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,852
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/310,357
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/102,865
FILING DATE: 06-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/604,075
FILING DATE: 18-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 91.201.523.7
FILING DATE: 18-Jun-1991
ATTORNEY/AGENT INFORMATION:
NAME: Mary E. Gormley
REGISTRATION NUMBER: 34,409
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2375 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
IMMEDIATE SOURCE:
CLONE: Em100E
FEATURE:
NAME/KEY: CDS
LOCATION: 3..1859
US-08-468-852-9

Query Match 8.6%; Score 27.2; DB 2; Length 2375;
Best Local Similarity 52.7%; Pred. No. 6.1;
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 47 aggttgatgggtgcttagctctcaaccatagacggtgcccgcacatctgctggggccaatt 106
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1576 AGATTGAAGCTGTGACAGGACGTACACAGAAAAACAAAGCAAGTATCTGGATCTAAGG 1635
QY 107 gcacgtcgtcgtggaagctctatatattggcgtcgtcgaccaccagagccatcat 158
|||| || || || || || || || || || || || || || || || || || || || || ||
Db 1636 TAGGGTGGCGCTGCTAGTGTGCTAGAGGTGGCTATGCTTATAAAGGAACCTCCTT 1687

RESULT 15
US-08-468-857-9
; Sequence 9, Application US/08468857
; Patent No. 5925347
; GENERAL INFORMATION:
; APPLICANT: van den BOOGAART, Paul
; APPLICANT: KOK, Jacobus Johannus
; APPLICANT: VERMEULEN, Arnoldus Nicolaas
; TITLE OF INVENTION: COCCIDIOSIS POULTRY VACCINE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5925347el Patent Department
; STREET: 1300 Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:40:30 ; Search time 320.8 Seconds
(without alignments)
480.118 Million cell updates/sec

Title: US-09-300-482-14

Perfect score: 410

Sequence: 1 cccacgcgcgcgcgtcatg.....gcaacggccgcgtcttgatg 410

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36:*

- 1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	99.2	24.2	1916	18	Brevibacterium fla
2	91.2	22.2	1038602	20	Complete genome se
3	69.2	16.9	14516	20	E. coli O111 antiq
4	69.2	16.9	14516	21	Escherichia coli f
5	66.2	16.1	4487	7	Sequence encoding
6	65.6	16.0	2450	20	Polynucleotide seq
7	57	13.9	2025	19	Streptococcus pneu
8	57	13.9	2025	19	Streptococcus pneu
9	56.4	13.8	1953	19	Streptococcus pneu
10	56.4	13.8	2087	20	Streptococcus pneu
11	49.6	12.1	16592	18	DNA encoding S. pn
12	39.2	9.6	53789	19	Staphylococcus aur
					Amycolatopsis medi

13	37.4	9.1	1524	18	T63571	Xylanase gene obta
14	35.6	8.7	939	20	X33560	Rice beta-glucanas
15	35.6	8.7	1020	20	X33570	Rice Gns8 CDS. Or
16	35.6	8.7	2088	12	Q14479	Truncated HSVgB ge
17	35.6	8.7	2612	20	X33536	Rice beta-glucanas
18	35.6	8.7	2925	12	Q14455	HSV surface antige
19	35.6	8.7	3465	12	Q14478	HSVgB gene. Herpe
20	35.6	8.7	8051	18	T2685	Sugar biosynthesis
21	35.6	8.7	8160	20	X25774	S.erythraea erythr
22	35.4	8.6	811	21	Z99478	Maize gibberellin
23	35.2	8.6	2298	20	X06781	Human testis secre
24	35.2	8.6	2990	19	V66832	zea mays soluble s
25	35.2	8.6	3476	18	T95698	Novel human gene,
26	34.8	8.5	1053	10	N90184	DNA encoding to 6-
27	34.8	8.5	1411	13	Q30672	Homo sapiens mamma
28	34.4	8.4	2277	19	V13836	Human telomerase p
29	34.4	8.4	2277	19	V05372	F. lutescens piper
30	34.2	8.3	6357	21	Z91051	Herpes Simplex Vir
31	34	8.3	2712	8	N71303	Glycoprotein B (gB
32	34	8.3	3472	14	Q48497	Herpes simplex vir
33	34	8.3	3473	18	T93651	Sequence of Herpes
34	34	8.3	3474	9	N80908	Sequence encoding
35	34	8.3	3966	6	N50516	Human retinal dege
36	34	8.3	5019	19	V34855	Streptomyces livid
37	33.8	8.2	1910	14	Q34549	Leishmania brazili
38	33.2	8.1	2040	18	T62137	Leishmania antigen
39	33.2	8.1	2040	19	V47557	Gene encoding a su
40	33.2	8.1	6741	21	A10595	A. crysogenum cyst
41	33	8.0	1778	16	Q86722	Sequence of herpes
42	33	8.0	2426	6	N50490	Herpes Simplex vir
43	33	8.0	2426	10	N90525	A. crysogenum cyst
44	33	8.0	3494	16	Q89311	Mycobacteriophage
45	33	8.0	49272	19	V35000	

ALIGNMENTS

RESULT	1
T88035	
ID	T88035 standard; DNA; 1916 BP.
XX	
AC	T88035;
XX	
DT	22-DEC-1997 (first entry)
XX	
DE	Brevibacterium flavum 6-phosphogluconate dehydrogenase DNA.
XX	
KW	Brevibacterium flavum; 6-phosphogluconate dehydrogenase;
KW	recombinant production; coryneform; bacterium; ds.
OS	Brevibacterium flavum.
XX	
FH	Key
FT	Location/Qualifiers
FT	374...1852
FT	/*tag= a
FT	/product= 6-phosphogluconate_dehydrogenase
PN	JP09224662-A.
XX	
PD	02-SEP-1997.
XX	
PF	23-FEB-1996; 96JP-0036346.
XX	
PR	23-FEB-1996; 96JP-0036346.
XX	
PA	(MITU) MITSUBISHI CHEM CORP.
XX	
DR	WPI; 1997-484097/45.
DR	P-PSDB; W27613.
XX	
PT	6-phospho:gluconate dehydrogenase - can be recombinantly produced by
PT	transforming coryneform bacteria with DNA molecule encoding it

Sat Nov 4 18:11:09 2000

```

XX PS Claim 1; Pages 5-7; 8pp; Japanese.
XX CC
XX CC The present sequence encodes the Brevibacterium flavum JM-233
XX CC 6-phosphogluconate dehydrogenase (6-PD). The 6-PD can be
XX CC recombinantly produced by transforming coryneform bacteria with the
XX CC DNA molecule encoding it.
XX PS Sequence 1916 BP; 425 A; 596 C; 507 G; 388 T; 0 other;

Query Match          24.2%; Score 99.2; DB 18; Length 1916;
Best Local Similarity 54.9%; Pred. No. 5.3e-17;
Matches 219; Conservative 0; Mismatches 174; Indels 6; Gaps 1;

QY 12 qcgggtcatggtggcgaacacttgccctcaacattgcagagaaagggttccccatctctgtg 71
Db 443 gcagtaatgggtcgaacactgcgcgaacttcgcccgaacaggccacactgtcgtgtc 502

QY 72 tacaacaggacaaacctccaaagggtggacgagaccgtgcagcgtgcccgaaggcagaagaaac 131
Db 503 tacaacgcgacgactgacaaacgcgaacgacagctcatcgcgcatcagcggctccgaaggcaac 562

QY 132 ctcccgctcagcgtctccatgaccgcgcgcgtcttctgtgaagtcattcagaagccaacgg 191
Db 563 ttcatcccttcgcaacccgtcgaaag-----ttcgtagcatccctggaaagccaacgc 616

QY 192 gtgggtgatcatgctcgtcaagcgcgcgcgcagtgccagttaccagaccatcggaagcctcgca 251
Db 617 cgcgcacatcatcattggttcaggctggttaacgcaccgcagcagtcacacacagctggca 676

QY 252 gtcacttggagcagggcagactgatacatcatcgtatggggggaacagagtgtacgaagaacacg 311
Db 677 gacgcacatggacaaaggcgacatcatcatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 736

QY 312 gagagaggagaaaggccatggagagcgcgcgcctnctgtatcttggcagtggtgtctct 371
Db 737 attctcgcggagagaaatctccgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 796

QY 372 ggaggaaggagggtggccgcgaacggccgcgcctcttgatg 410
Db 797 ggcgcgaagaaggcgactcaacggccacatccatcatg 835

RESULT 2
Z01425/c
ID Z01425 standard; DNA; 1038602 BP.
XX AC
XX AC Z01425;
XX DT
XX DT 07-OCT-1999 (first entry)
XX DE
XX DE Complete genome sequence of Chlamydia trachomatis.
XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
XX KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
XX KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX KW bartholinitis; pneumonia; venereal lymphogranulomatosis; ss.
XX OS
XX OS Chlamydia trachomatis.
XX PN
XX PN W09928475-A2.
XX PD
XX PD 10-JUN-1999.
XX PF
XX PF 27-NOV-1998; 98WO-TB01939.
XX PR
XX PR 04-NOV-1998; 98US-0107077.
XX PR 28-NOV-1997; 97FR-0015041.
XX PR 17-DEC-1997; 97FR-0016034.
XX PA
XX PA (GEST ) GENSET.

```

```

PI Griffiths R;
XX DR WPI; 1999-371125/31.
XX PT Genome sequence of Chlamydia trachomatis
XX PS Claim 1; Page 373-656; 1755pp; English.
XX CC The present sequence represents the complete genome of Chlamydia
XX CC trachomatis. Open reading frames (ORFs) of the genome encode polypeptides
XX CC Y3754-Y37949. The polypeptides can be used as vaccines against
XX CC Chlamydia trachomatis. Antisense and ribozyme sequences can also be
XX CC used to control growth of the microorganism. Chlamydia trachomatis is
XX CC responsible for a large number of diseases, e.g. eye diseases such as
XX CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
XX CC conjunctivitis; genital diseases such as nongonococcal urethritis,
XX CC epididymitis, cervicitis, salpingitis, perihhepatitis, bartholinitis;
XX CC pneumonia in breast feeding infants; and venereal lymphogranulomatosis.
XX CC The polypeptides of the invention may be of use in treating these
XX CC diseases.
XX PS Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;
XX QY

Query Match          22.2%; Score 91.2; DB 20; Length 1038602;
Best Local Similarity 53.6%; Pred. No. 3.3e-14;
Matches 214; Conservative 0; Mismatches 179; Indels 6; Gaps 1;

QY 12 qcgggtcatggtggcgaacacttgccctcaacattgcagagaaagggttccccatctctgtg 71
Db 397456 GCGGTGATGGGCAAAAACCTTGTATTGAACATGGTGGATCATGGTGTTCCTTTCTCTCTGTC 397397

QY 72 tacaacagacaaacctccaaagggtggacgagaccgtgcagcgtgcccgaaggcagaagaaac 131
Db 397396 TATAACCGAAGTCCCGCGGAAACACAGAGAGTCTTGAAGATCATGGAGAGTGGAGCT 397337

QY 132 ctcccgctcagcgtctccatgaccgcgcgcgcgtcttctgtgaagtcattcagaagccaacgg 191
Db 397336 CTGC-----AAGGATTTACTACGATTCAGAGTTCCTTGAAGCGTCTCTCGT 397283

QY 192 gtgggtatcatgctcgtcaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 251
Db 397282 AAGATCATGATCATGATTAAGCGGAGCTCCCTGTTGTAAGATGATTCCTCCCTGCTT 397223

QY 252 gctcacttgagcagggcgactgcatactcagtgagggggaacgagtggtacgagaacacg 311
Db 397222 CCTTCTTGGAGAGGAGAGATATTCATTGATGGGGGAATAGCTATTATTAGATTCT 397163

QY 312 gagagaggaggagagccatggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 371
Db 397162 GAGCGACGCTATGTCGACCTGAAAAAAGAGAAATTCATTCTTGGATGGGAGTCTCT 397103

QY 372 ggaggaaggagggtggccgcgaacggccgcgcctcttgatg 410
Db 397102 GGAGGGGAAGAGGGGGCTAGAAAAAGGCGCTTCCATTATG 397064

RESULT 3
X06748
ID X06748 standard; DNA; 14516 BP.
XX AC
XX AC X06748;
XX DT
XX DT 26-APR-1999 (first entry)
XX DE
XX DE E. coli O111 antigen gene cluster.
XX KW O antigen; O111 antigen; diagnosis; wbdH gene; wzx gene; wzy gene;
XX KW wbdM gene; rfb gene; transferase; polymerase; flippase; diarrhoea;
XX KW haemorrhagic colitis; ds.
XX OS
XX OS Escherichia coli.
XX PA

```


PA (UNSY) UNIV SYDNEY.

XX Reeves PR, Wang L;

XX WPI; 2000-072598/06.

XX Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc. -

XX Claim 3; Page 216-221; 245pp; English.

XX Z56331 to Z56398 represent nucleic acid molecules (I) encoding all or
CC part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (1) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (I) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H-serotypes of E. coli by hybridising at least one or a pair of (I)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (1) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (I) when using at least one
CC (I) is detected by southern blot analysis and, when using a pair of (I),
CC is detected by polymerase chain reaction (PCR). Z56399 to Z56420
CC represent primers used in the exemplification of the present invention.

XX Sequence 14516 BP; 4582 A; 2221 C; 2971 G; 4742 T; 0 other;

Query Match 16.9%; Score 69.2; DB 21; Length 14516;
Best Local Similarity 50.4%; Pred. No. 5.7e-09;
Matches 201; Conservative 0; Mismatches 189; Indels 9; Gaps 1;

QY 12 ggggtcattggggcagaaacttgccctcaacattgcagagaaagggttcccatctctgtg 71

Db 13157 gcaagtggggcgaacactgacgtgcgtcaacatcgaagccggttatccgtctccatc 13216

QY 72 tacaacagacaacctccaaggtgacagacacgtgcagcgtgccagggcagaaggaaac 131

Db 13217 tcaacccctcccgagaaactgaagaagtgtgtgccgaacccggataaagaactg 13276

QY 132 cttcccgctcattcgttccatgaccccgctctcttggtaagtccattcagaagccacgg 191

Db 13277 gttccctattacacggtgaaga-----gttcgtcgagctcttgaacccacgt 13327

QY 192 gtggtgatcatgctcgtcaaggccgcccagtcagcagtgaccagacatcgcagcctcgca 251

Db 13328 cgtatcctgttaattgttaaaagcagggcggaactgatgtcgtatcgtatccctgaag 13387

QY 252 gctcacttgagcagggcactgcacatcagatgggggaaacagagtggtacgaaacacg 311

Db 13388 ccgtatcttgataaaggagacatcattatgatgtggcacaacaccttctccagacact 13447

QY 312 gagagagggagaaagcccatgagagcgcggcctnctgtatcttggcatgggtgtctct 371

Db 13448 atccgctgaacgfgaactgccgcggaaggccttaacttcatcgttacccggcgtgtcc 13507

QY 372 ggaggaagaggggtgccgcgaacggcccgctcttgatg 410

Db 13508 ggcggtgaagaggcgccctgtaagggcccatctatcatg 13546

RESULT 5

N60943

ID N60943 standard; DNA; 4487 BP.

XX

AC N60943;

XX 07-OCT-1991 (first entry)

XX Sequence encoding gluconic acid promoter and operon.

XX Catabolite repression; ss.

XX Bacillus subtilis.

XX JP61100194-A.

XX 19-MAY-1986.

XX 23-OCT-1984; 84JP-0221192.

XX 23-OCT-1984; 84JP-0221192.

XX (MITK) MITSUI TOATSU CHEM INC.

XX WPI; 1986-166561/26.

XX Gluconic acid operon and its promoter of Bacillus subtilis - in
PT DNA fragment, which is controlled by catabolite repression.

XX Claim 2; Page 499; 10pp; Japanese.

XX The sequence may be used in the expression of hetrogenous sequences,
CC without catabolite repression, inducible with the action of gluconic
CC acid.

XX Sequence 4487 BP; 1308 A; 991 C; 1065 G; 1123 T; 0 other;

Query Match 16.1%; Score 66.2; DB 7; Length 4487;

Best Local Similarity 50.0%; Pred. No. 2.6e-08;

Matches 198; Conservative 0; Mismatches 189; Indels 9; Gaps 1;

QY 15 gtcatggggcagaaaccttgcctcaacattgcagagaaagggttcccatctctgttac 74

Db 3932 gtaatgggaagcaacatcgcttaacattgcgaataaaggcgaaacgtcgctgtctat 3991

QY 75 aacagacaaacctccaaggtgacagacacgtgcagcgtgccagggcagaagdaaacctt 134

Db 3992 aa-----ttaccaccagagatttaacggaccagcttatccaaaagctggtgacaa 4042

QY 135 ccgctctacggcttccatgaccccgctctcttggtaagtcctcattcagaagccacgggtg 194

Db 4043 tctctcagcccgctattacagcgttgaggactttgttcaatcgttagaaaaaaccaagaaa 4102

QY 195 gtgatcatgctcgtcaaggccgcccagtcgacagaccatcgcagcgtcgcagct 254

Db 4103 atcttttggatggtcagcagcgggaaacccgtagattccgtcatccaatcattaaagct 4162

QY 255 caettgagcagggcagcactgcacatcagatgggggaaacagagtggtacgagaacacggag 314

Db 4163 ttgcttgaaagagcagcgtcatcattgagcggagggaaccccccactcatgaagacagaa 4222

QY 315 aggaagggagagggccatggagagcgcggcctnctgtatcttggcatgggtgtctctgga 374

Db 4223 agaagatatcagcagctgaaggaagaaagggatcgcttaoctgggagtcggtcattccggc 4282

QY 375 ggaagaggggtgccgcgaacggcccgctcttgatg 410

Db 4283 ggtgaagtcggtcggttaacagggcccttccatcatg 4318

RESULT 6

X20537

ID X20537 standard; DNA; 2450 BP.

XX

AC X20537;

XX 05-MAY-1999 (first entry)

DT

Db 232 cacaccttgacagggtgatcttcttgatgacggtgaaatacttcttccaaagatacca 291
 Qy 313 agaggaggagagccatgagagagcggcgccctnctgtatcttggcagtggtctctg 372
 Db 292 tccgtcgtaataagaattggcaactcaggtatcaactttatcgtactgaggtttctg 351
 Qy 373 gagaaagagggtgcccgcacacgcccgtcccttgatg 410
 Db 352 gtggtgaaaaaggtgccccttgagggtcccttctatcatg 389
 RESULT 8
 X16683
 ID X16683 standard; DNA; 2025 BP.
 AC X16683;
 XX
 DT 04-MAY-1999 (first entry)
 XX Streptococcus pneumoniae response regulator ORF DNA sequence.
 DE
 XX Streptococcus pneumoniae; response regulator; antibacterial; infection;
 KW pneumonia; bacteraemia; meningitis; otitis media; conjunctivitis;
 KW sinusitis; pleural empyema; endocarditis; gene therapy; ss.
 XX Streptococcus pneumoniae.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1403..2023
 FT /*tag= a
 FT /note= "no stop codon given"
 XX
 XX EP892057-A2.
 XX
 XX 20-JAN-1999.
 XX
 XX 10-JUL-1998; 98EP-0305517.
 XX
 XX 18-JUL-1997; 97US-0053238.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 XX Biswas S, Kosmatka AL, Shilling LK, Throup J, Wallis NG;
 XX Zalacain M;
 XX
 XX WPI; 1999-083574/08.
 XX P-PSDB; W94667.
 XX
 XX New Streptococcus pneumoniae response regulator polypeptides and
 XX polynucleotides - useful as diagnostic reagents and for prevention
 XX and treatment of Streptococcus pneumoniae infections, especially
 XX pneumonia, bacteraemia and meningitis

Claim 2; Page 5-6; 43pp; English.

The present sequence encodes a bacterial response regulator (RR) protein which is a component of the two component signal transduction system (TCSTS). RR polynucleotides and proteins are useful for diagnosing susceptibility to diseases by detecting mutations or polymorphisms in the RR gene or analysing for the presence of amount of RR protein expressed in a patient sample. RR PCR probes are useful for diagnosing diseases, and can characterise the response of the infectious organism to drugs. RR proteins and polynucleotides are also useful for screening for antagonists, agonists and drugs against infectious micro-organisms. RR agonists and antagonists are bacteriostatic and bacteriocidal compounds which can be used in treatment to enhance (agonist) or block (antagonist or antisense sequence) RR activity, therefore treating microbial diseases, especially Streptococcus pneumoniae diseases including otitis media, bacteraemia, conjunctivitis, pneumonia, sinusitis, pleural empyema, endocarditis and especially meningitis. Epitopes of RR proteins and polynucleotides are useful immunogens for producing anti-RR antibodies for prevention of bacterial infections, and

CC RR polynucleotides can be used in genetic immunisation (gene therapy) to prevent infections. RR proteins, polynucleotides and their (ant)agonists can prevent adhesion of bacteria to matrix proteins, and are useful for use on wounds and body implants to prevent bacterial infection.
 XX
 SQ Sequence 2025 BP; 592 A; 413 C; 463 G; 557 T; 0 other;

Query Match 13.9%; Score 57; DB 20; Length 2025;
 Best Local Similarity 48.5%; Pred. No. 5.2e-06;
 Matches 193; Conservative 0; Mismatches 196; Indels 9; Gaps 1;

Qy 13 cggctcatgggagcagaacctccctcaacattgcagagaaaggggtcccccctctctgtg 72
 Db 1 cggtaattgggtcgtaaccttgccttaattgaatcaatcagtggttacacaattgtatct 60
 Qy 73 acaacaggacacacctcccaagtgagcagccgtgcagcgtgcacagggcagaaggaacc 132
 Db 61 acaaccgtagtaagaaaaaacggaagatgtgattgcttgcctcctgaaaagaactttg 120
 Qy 133 ttccogtctacggttccatgaaccccgctctcttggtaagtccattcagaagccacggg 192
 Db 121 taccagct-----atgacgttgaaagtttgaactcaatcgaataaaacctcgtc 171
 Qy 193 tgggtatcatgctcgtcaagcgcgcgcaggttgacacagaccatcgcgacgtctcgag 252
 Db 172 gtatcatgctgattgttcaagctggaacctgtacagatgctactatccaagcccttctc 231
 Qy 253 ctcaattggagcagggcgactgcatactcgtatggggggaacagagtggtacagaaacagg 312
 Db 232 cacaccttgacaagggtgatctcttgattgacggtggaaatactttctacaagatacca 291
 Qy 313 agagaggaggagaagcccatggagagcgccctnctgtatcttggcatgggtgtctctg 372
 Db 292 tccgtcgtaatgaagaattggcaactcaggtatcaactttatcgggtactggagttctg 351
 Qy 373 gaggaaggagggtgcccgcacacggccgctctcttgatg 410
 Db 352 gtggtgaaaaaggtgccccttgaaaggtccttctatcatg 389

RESULT 9

V52287
 ID V52287 standard; DNA; 1953 BP.
 XX
 AC V52287;
 XX
 XX 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:154.
 XX
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 XX WO9818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 XX WPI; 1998-272225/24.
 XX
 XX Computer-readable medium with recorded Streptococcus pneumoniae
 XX polynucleotide sequences - useful in diagnostic kits and assays, and

pharmaceutical compositions and vaccines for Streptococcus pneumoniae

Claim 1; Page 1020-1021; 1409pp; English.

The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (V52134 to V52524) are genomic fragments from *Streptococcus pneumoniae*. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the *S.pneumoniae* genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridise to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the *S. pneumoniae* genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the *S. pneumoniae* genome of commercial importance, or expression modulating fragments of the *S. pneumoniae* genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for *S. pneumoniae*.

Sequence 1953 BP; 568 A; 404 C; 443 G; 538 T; 0 other;

Query Match	13.8%;	Score 56.4;	DB 19;	Length 1953;
Best Local Similarity	48.4%;	Pred. No. 7.3e-06;		
Matches 193;	Conservative	0;	Mismatches 197;	Indels 9; Gaps 1;

12	gcggtcatggggcagaaaccttgcctccaacattgcagagaaagggtcccccattctgtg	71
13		
14		
44	gcctgaatgggtcgttaaccttgccttaattgaatctggttacacagttgctatc	103
45		
46		
72	tacaacaggcaaaccttccaagtggaacagaccgtgcagcgtgccaaaggcagaaggaac	131
73		
74		
104	tacaacgctagtataagaaaaacggaaagtgtattgtcttgcctatcctgaagaacattt	163
105		
106		
132	cttcccgtctacggcttccatgacccccgcctcttctgaagtcatttcagaagccaagg	191
133		
164	gtaccaagct-----atgacgttgaagaatttgttaacctcaatcgaagaaacctctg	214
165		
192	gtgggtatcatgctctcaaggccgcgcgcagttgaccagaccatcgcagcgtcgca	251
193		
215	cgtatcatgctgattgttcaagctggacctggtacagatgctactatccaagccctctt	274
216		
252	gctcaacttggagcaggcgactgcatactatcgtgggggaacagagtggtacagaaacag	311
253		
275	ccacaccttgacaagggtgatctctgattgacggaggaataacttcttacaagaatacc	334
276		
312	gagagaggaggagaagcccatggaggcgcgcgcctnctgtatcttggcatgggtgctct	371
313		
335	atccgctcgaattgaagaattgcgaaccttggtatcaactttatcgttgggtctctt	394
336		
372	ggaggaaaaaggaggttgcccgcaacggccgttccttgatg	410
373		
395	ggctgtgtaaaaaggctgcccttgaaggtccttctatcatc	433
396		

RESULT 10
330047
D 230047 standard; DNA; 2087 BP.
X
X
C 230047;
X
X
25-JAN-2000 (first entry)
X
X
E DNA encoding *S. pneumoniae* 6-phosphogluconate dehydrogenase protein.

XX KW KW KW YV

XX
OS
Streptococcus pneumoniae.

FH	Key	Location/Qualifiers
100	100	100
101	101	101
102	102	102
103	103	103
104	104	104
105	105	105
106	106	106
107	107	107
108	108	108
109	109	109
110	110	110
111	111	111
112	112	112
113	113	113
114	114	114
115	115	115
116	116	116
117	117	117
118	118	118
119	119	119
120	120	120
121	121	121
122	122	122
123	123	123
124	124	124
125	125	125
126	126	126
127	127	127
128	128	128
129	129	129
130	130	130
131	131	131
132	132	132
133	133	133
134	134	134
135	135	135
136	136	136
137	137	137
138	138	138
139	139	139
140	140	140
141	141	141
142	142	142
143	143	143
144	144	144
145	145	145
146	146	146
147	147	147
148	148	148
149	149	149
150	150	150
151	151	151
152	152	152
153	153	153
154	154	154
155	155	155
156	156	156
157	157	157
158	158	158
159	159	159
160	160	160
161	161	161
162	162	162
163	163	163
164	164	164
165	165	165
166	166	166
167	167	167
168	168	168
169	169	169
170	170	170
171	171	171
172	172	172
173	173	173
174	174	174
175	175	175
176	176	176
177	177	177
178	178	178
179	179	179
180	180	180
181	181	181
182	182	182
183	183	183
184	184	184
185	185	185
186	186	186
187	187	187
188	188	188
189	189	189
190	190	190
191	191	191
192	192	192
193	193	193
194	194	194
195	195	195
196	196	196
197	197	197
198	198	198
199	199	199

FT	CDS	11..1435

```

EIT /*tag= a
EIT /product= "6-phosphogluconate dehydrogenase"

```

XX
DN
W00052020-11

PN WO9953020-A1.

XX
PD 21-OCT-1999

XX

09-APR-1999; 99WO-US07782.
PF XX

PR 10-APR-1998; 98US-0058692.
YY

PA (SMIK) SM

XX
PT
Warren Pt.:[illegible]

DR
WPI; 1999-620409/53.
P-PSDB; V43524.
XX
PT New 6-phosphogluconate dehydrogenase from *Streptococcus pneumoniae*, and
PT related proteins, used to screen for antibacterial agents, to treat
PT infections and to raise diagnostic antibodies -

XX
PS Claim 2: Page 3-4: 62pp: English

The present sequence encodes a 6-phosphogluconate dehydrogenase protein from *Streptococcus pneumoniae*. The protein induces an immune response against bacteria, causing production of antibodies that interfere with bacterial adhesion. The 6-phosphogluconate dehydrogenase, its fusion proteins and cells (or their membranes) are used to screen for specific modulators of its function or expression, which are potentially useful as antibacterials and to raise antibodies. Antibodies are used as immunoassay reagents; for isolation or identification of 6-phosphogluconate-expressing clones; for affinity purification; and as therapeutic antibacterials. Conditions that can be treated include bacterial infections, by *Helicobacter pylori*, or especially by *S. pneumoniae*, particularly meningitis. Detecting the presence of the protein or mutations in gene sequences encoding it can be used for the diagnosis, staging or monitoring treatment of infections, or determining susceptibility to them, particularly in standard immunoassays or hybridisation and amplification tests. Nucleic acids encoding the protein may also be used for chromosomal mapping and identification, for genetic immunisation or the rational design of (ant)agonists.

Sequence 2087 BP: 610 A: 424 C: 474 G: 572 T: 7 other:

Query Match	13.8%	Score 56.4;	DB 20;	Length 2087;
Best Local Similarity	48.4%	Pred. No. 7.5e-06;		

12 gcgctcatgggcagaaaccttccctcaacattgcagaagaagggtccccatctctgtg 71
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
bb 44 gcgtaattgggtgaacttcgcccttatattgaatcwtggttacacagtttgtatc 103
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

72 tacaacagcagacaacctccaagtgcagagaccctgcagcgtgcacaaggcagaaggaac 131
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
bb 104 tacaaccgtagtaagaaaaaacggaagatgatgtcttgcatactgaaaagaacctt 163
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

132 ctcccgtctaogcttccatgaccccgctcttttgtaagtcattcagaagcacgg 191
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
bb 164 gtaccaagct-----atgaogttgaaagttttgtaacctcaatcgaaaaacctgt 214
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

192 gtggtgaatatcgtcgtcaaagcgcgccagttgaccagaccattcgcgacctcgca 251
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||


```

RESULT 15
X33570/c
ID X33570 standard; DNA; 1020 BP.
XX
XX AC X33570;
XX DT 07-JUL-1999 (first entry)
XX DE Rice Gns8 CDS.
XX KW Rice; beta-glucanase; Gns; promoter; plant resistance; expression;
XX KW fungal infection; transgenic monocotyledon; growth; ss.
XX OS Oryza sativa.
XX PN WO9859046-Al.
XX PD 30-DEC-1998.
XX PF 25-JUN-1998; 98WO-US13525.
XX PR 25-JUN-1997; 97US-0050675.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Rodriguez RL;
XX DR WPI; 1999-105620/09.
XX DR P-PSDB; Y09305.
XX
XX New polynucleotides which hybridise with rice beta-glucanase genes -
XX useful for transforming monocot plants for various characteristics
XX including increased resistance to fungal infection
XX
XX Disclosure; Page 74-75; 90pp; English.
XX
XX The present invention describes isolated polynucleotides (PN) comprising
XX a sequence which hybridises under high stringency with a rice
XX beta-glucanase gene. Also described are: (A) a chimeric gene (CGL) used
XX to produce a transgenic monocot plant, comprising: (i) a transcriptional
XX regulatory region which hybridises under high stringency with a rice beta
XX -glucanase promoter; and (ii) a DNA sequence heterologous to the
XX regulatory region, and encoding a protein to be produced by the plant;
XX and (iii) a second DNA sequence encoding a signal polypeptide operatively
XX linked 5' to 3', so that the signal polypeptide is in translation frame
XX with the protein, and is effective to facilitate secretion of the protein
XX across aleurone or scutellar epithelium layers into the endosperm of
XX seeds obtained from the plant; (B) a monocot plant stably transformed
XX with CGL; (C) seeds from the above plant; (D) a method of enhancing the
XX resistance of a monocot plant to fungal infection by stably transforming
XX it with CGL; (E) a method of producing a heterologous protein, comprising
XX stably transforming a plant with CGL, obtaining seeds from the
XX transformed plant, germinating the seeds and obtaining the protein from
XX the seed endosperm; (F) an isolated protein having the characteristics of
XX a rice beta-glucanase enzyme, corresponding to beta-glucanases 2-9
XX (Gns 2-9). The polynucleotides of the invention are useful to transform
XX monocot plants. The polynucleotides of the invention are useful to transform
XX provide increased resistance to fungal infection, improved growth
XX characteristics, and high expression levels of heterologous proteins in
XX various tissues obtained from the plant.
XX
XX Sequence 1020 BP; 172 A; 360 C; 320 G; 168 T; 0 other;

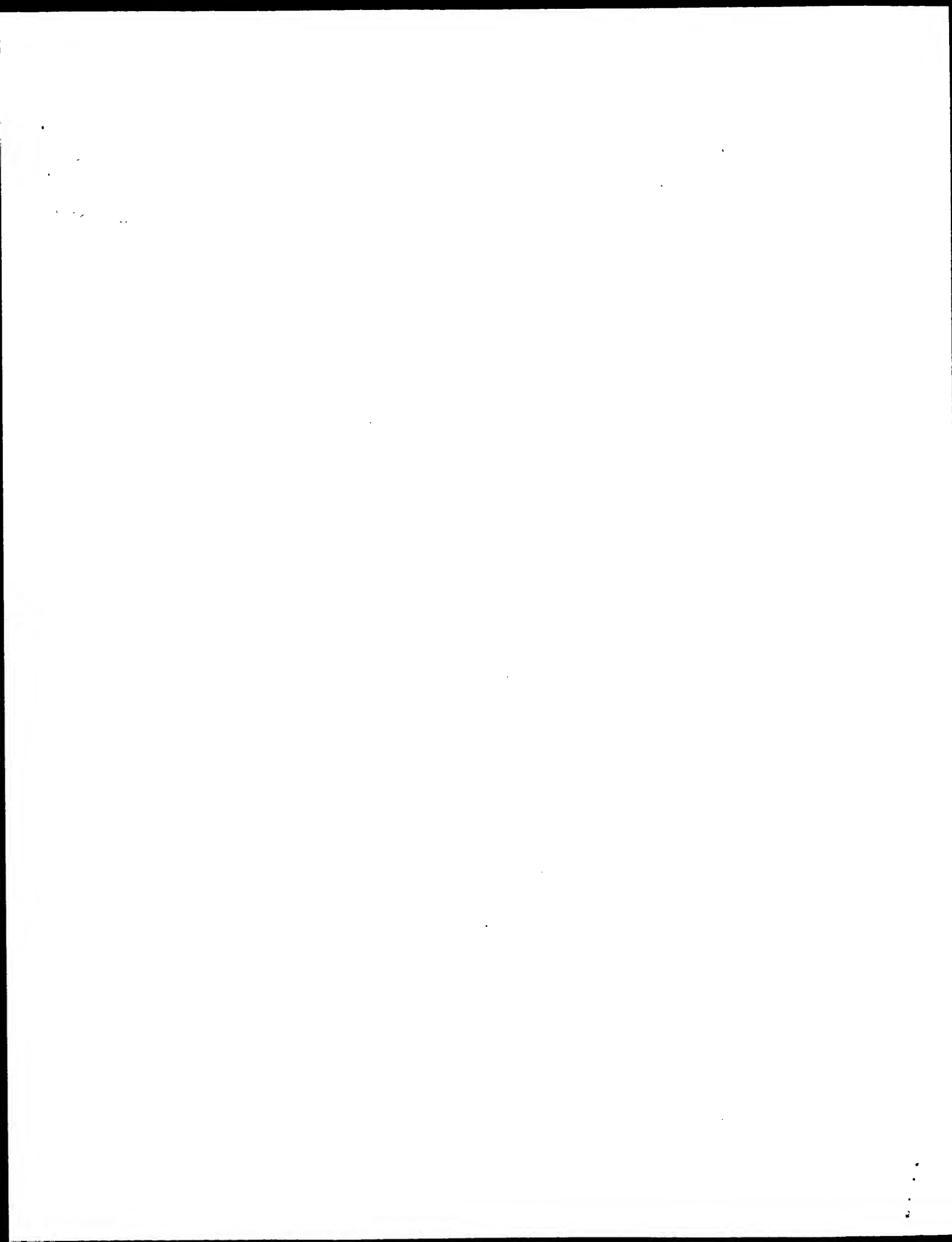
```

Search completed: November 4, 2000, 13:43:19
Job time: 16326 sec

```

Query Match      8.7%; Score 35.6; DB 20; Length 1020;
Best Local Similarity 45.9%; Pred. No. 1.6;
Matches 119; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 92 ggttgagacagaccgtcagcggtccaaaggcagagaaacattccctcgttaaggttcca 151
Db 702 GGTGTATGTACCGCGCGTCTGTCACCGAGCCGACTGCGACCCCGAGCAGCAGCGCGTA 643

```




```
; ANTI-SENSE: no
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM:
; FEATURE:
; NAME/KEY: sequence of xylanase gene identified by
; NAME/KEY: amplification of xylanase fragments from soil
US-08-716-942-24

Query Match          9.1%; Score 37.4; DB 3; Length 1524;
Best Local Similarity 50.6%; Pred. No. 0.2;
Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 189 cgggtggtgatcgtctcaaggcgccgagccagtggtgacccagaccatcgagcgtc 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 595 CGCGGATGGGACGCGGTATACGCGGAGCGGCTGCGCACCGCGCACGCGCGGACCC 654

Qy 249 gcagctcaactggagcggcgactgcatcgatggggggaacgagtggtacgagaac 308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 655 AGTCCCGAGCTGTACATCAAGGACTACAGCATCGAGGCGGAGAGACGCCAAGACGACGCC 714

Qy 309 acgagagaggaggaagccatgagagcgccgctnctgtatcttggcatggg 364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 715 ATGTACAGCTGGTGCAGTCCCTGCTGCGCAGGCGGTGCGGATCAACAGCGGTGGG 770

RESULT 2
US-08-576-626A-2
; Sequence 2, Application US/08576626A
; Patent No. 598194
; GENERAL INFORMATION:
; APPLICANT: Summers, R.G.
; APPLICANT: Katz, L.
; APPLICANT: Donadio, S.
; APPLICANT: Staver, M.J.
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
; TITLE OF INVENTION: BIOSYNTHESIS GENES
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/576.626A
; FILING DATE: 21-DEC-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dianne Casuto
; REGISTRATION NUMBER: P-40,943
; REFERENCE/DOCKET NUMBER: 5857.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 938-3137
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8051 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-576-626A-2

Query Match          8.7%; Score 35.6; DB 4; Length 8051;
Best Local Similarity 45.6%; Pred. No. 1;
Matches 125; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 71 gtacacagcaggaacactccaagggtgacgagaccgtgcagcgtgccaaggcagaagaaa 130
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2317 GTCCGACAGCACACTCCGGAGGGGAGCAGGATGTACAGAGGGGGTTCGCCGAGCITTA 2376

Qy 131 cettcccgtctacggcttcctaccgcccgctcctttgtgaagtccattcagaagccacg 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2377 CGACCGGTTTACCGCGCGCGGCGCAAGGACTACGCGCGCGGAGCGCGCAGTCCGCGC 2436

Qy 191 ggtgtgatcgtctcgtaaggccgagccgagccgagccagttgacagaccatcgagcgtcgc 250
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2437 GGTGTGTCAGAGACCGCGCTCGCGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2496

Qy 251 agctcacttggagcagggcgactgcatcgatgggggaacgagtggtgtacgagaacac 310
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2497 CACCCACCTGCGCGGTTGCGCGGACCTCTTCGACGACGTGACCGGGCTGGAGCTGTCG 2556

Qy 311 ggagaggggggagagggccatggagggagcgcgcgc 344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2557 GCGGATGATCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2590

RESULT 3
US-08-572-951-1/c
; Sequence 1, Application US/08572951
; Patent No. 5824790
; GENERAL INFORMATION:
; APPLICANT: KEELING, PETER L.
; APPLICANT: KNIGHT, MARY E.
; APPLICANT: GUAN, HANPING
; TITLE OF INVENTION: MODIFICATION OF STARCH
; TITLE OF INVENTION: SYNTHESIS IN PLANTS
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; ADDRESSEE: Intellectual Property Group of
; ADDRESSEE: Pillsbury Madison & Sutro LLP
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/572,951
; FILING DATE: 15-DEC-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/346,602
; FILING DATE: 29-NOV-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/263,921
; FILING DATE: 21-JUN-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul N. Kokulis
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 222957/1.02.15C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```


Db	960	GGACCTGGTGTGTCAGGTGGACCTATGGAACCAAGAGGTGTAGAACTGTGGAGCCGGA	1019
Qy	239	cgcagcgtcgagctcaacttgagcagggcgactgcatcatcgatgggggggaacgagtg	298
Db	1020	CCAGGAGGTTACGACCCTGGTGAGCCCGAACCAAGAGGTTATGACACTGTGTGGAGCCGGA	1079
Qy	299	gtacgagaacaaggagagggagaaagccatggagagcgcgcctnctgtatttgg	358
Db	1080	CCAGGAGGTTACGACCTGGTGGACCTGGACCAAGAGGTACCGACCTGTGTGGAGCTGGG	1139
Qy	359	catgggtgtctctggagaaaagggtgtcccgcaacggccccttcgtgatg	410
Db	1140	CCTGGAGGTTACGAGCCTGGTAGCTGGACCTGCAGCTGCAGGTTACGACCACTGGTG	1191

```

RESULT      6
US-08-676-967-5
; Sequence 5, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-676-967-5

```

[illegible]

```

Db      681 CATGGAGGAGGAGAACCGACGACGACGACGACGACGACGAGGAGGAGCGCGGT 740
QY      281 cgaTggggggaacgagtggtacgagaaacacggagaggaggaga 324
Db      741 CGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 784

RESULT 7
US-08-676-974-5
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-676-974-5

```

[illegible]

CITY:	Palo Alto
STATE:	CA
COUNTRY:	USA
ZIP:	94304-1018
COMPUTER READABLE FORM:	
MEDIUM TYPE:	Floppy disk
OPERATING SYSTEM:	IBM PC compatible
SOFTWARE:	Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:	
APPLICATION NUMBER:	US/08/720,229
FILING DATE:	26-SEP-1996
CLASSIFICATION:	424
ATTORNEY/AGENT INFORMATION:	
NAME:	Schiff, J. Michael
REGISTRATION NUMBER:	40,253
REFERENCE/DOCKET NUMBER:	29938-20002.00
TELEPHONE:	(415) 813-5600
TELEFAX:	(415) 494-0792
TELEX:	706141
INFORMATION FOR SEQ ID NO:	13:
SEQUENCE CHARACTERISTICS:	
LENGTH:	2713 base pairs
TYPE:	nucleic acid
STRANDEDNESS:	double
TOPOLOGY:	linear
US-08-720-229-13	

Query Match	8.3%	Score 34;	DB 5;	Length 2713;
Best Local Similarity	47.2%	Pred. No. 2;		
Matches 103; Conservative	0;	Mismatches 115; Indels	0;	Gaps

QY	74	caacaggacaacctccaagtgtagcagaccgtgcagcgctgccaaggcagaaggaacct	133
Db	1482	CAGACCACTCCTCCATGAGTTCGCCGGGTGCAGTTAGTGTAACCAACCATACAGCG	1541
QY	134	tccgtctacggcttcattgaccccgcgtccttttgaagtccattcagaagccacgggt	193
Db	1542	CCATGTCACAGATATGTTGGCGCGTTCCTATCCGTTGCGAGCTGCAGAATCACGA	1601
QY	194	ggtgatcatgctgtcaagccgccccgcagctgaccagaccatcgagcgtcgacgc	253
Db	1602	GCATGACCTGTGAACGAGCCCGCAAGTGRACCCACGCGATCGCTCGCCACCGT	1661
QY	254	tcacttggagcaggcgactgcatcatcgatgggggga	291
Db	1662	GGCGCGGGGTGAGCGCGGATGTCGCGACGTGA	1699

RESULT 11	
5244792-2	
Patent No.	5244792
APPLICANT:	BURKE, RAE L.; PACHL, CAROL VALENZUELA, PABLO D.T.
TITLE OF INVENTION:	EXPRESSION OF RECOMBINANT GLYOPROTEIN
FROM HERPES SIMPLEX VIRUS	
NUMBER OF SEQUENCES:	19
CURRENT APPLICATION DATA:	
APPLICATION NUMBER:	US/08/587,179.
FILING DATE:	20-SEP-1990
PRIOR APPLICATION DATA:	
APPLICATION NUMBER:	921,730
FILING DATE:	20-OCT-1986
APPLICATION NUMBER:	597,784
FILING DATE:	06-APR-1984
SEQ ID NO:2:	
LENGTH:	3472
5244792-2	

Query Match	8.3%	Score 34;	DB 7;	Length 3472;
Best Local Similarity	47.2%	Pred. No. 2;		
Matches 146; Conservative	0;	Mismatches 165; Indels	3;	Gaps

QY	12	gcggctatggggcagaaccttgcctcaaatgcagagaaaggggtccocatctctgtg	71
----	----	---	----

```

Db 1108 GAGTTCTTCGATGCGACGACGAGACCTGTCCGACACAGAGAAATGTTCCCAAGGACATC 1167
QY 72 tacaacagga---caacctccaagtgaacgagacggtgcagcgtgccaagcgagaaagga 128
Db 1168 ACCAAGTGAGGCTCAATGACCTCATGGACAAGATCGAGAGCCCGAGCCGGAAGACACA 1227
QY 129 aacctcccggtacggttccatgaccgccggtccttgtgaagtccattcagaagcca 188
Db 1228 CAAGATGCTCTACCGCGCAGGGTCCCTGAGTTCAAGGTGGCTCCAGTGTGAGGAG 1287
QY 189 cgggtggtgatcatctcgtcaagcccgccgagccagtgtagccagacatcgacgctc 248
Db 1288 CTGAACATCATAGAGGACGAGGTTAGCCAGCGCTGTGCTGACCGCCCTCCAAAGATCCAC 1347
QY 249 gcagctcacttgagcagggcgactgcacatcatcgatgggggagcagtggttaagagaac 308
Db 1348 GTGCTGCTATTGGTCTGCACGAGGACCACTCTCTGGACACAGGCGCGGGGACCCCGCAGC 1407
QY 309 acggagagaggagga 322
Db 1408 TCCAAGAAGGGCGA 1421

```

RESULT 13

```

PCT-US92-05532-1/c
; Sequence 1, Application PC/TUS9205532
; GENERAL INFORMATION:
; APPLICANT: Lichenstein, Henri
; TITLE OF INVENTION: Isolation and
; TITLE OF INVENTION: Characterization of a Novel Protease
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, DS, 1.4 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh OS 6.0.4
; SOFTWARE: Microsoft Word Version 4.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/05532
; FILING DATE: 19920629
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1910 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double stranded
; TOPOLOGY: circular
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: DNA sequence encoding
; NAME/KEY: Streptomyces lividans protease x
; OTHER INFORMATION: bp 12 to 1910 comprises
; OTHER INFORMATION: S. lividans DNA sequence, including Protease X.
; OTHER INFORMATION: bp 1 to 11 represents cloning vector DNA sequence.
PCT-US92-05532-1

```

```

Query Match      8.2%; Score 33.8; DB 6; Length 1910;
Best Local Similarity 47.8%; Pred. No. 2;
Matches 98; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 96 gacgagaccgtgcagcgtgcgaagcgagaggaacaccttcccgctctacggcttccatgac 155
Db 551 GCCAACGCGCTGTCGCGCGGGTGACCGAGGTCAGCTGACGCTGCGCTTCGTGCGACGAG 492

```

```

QY 156 cccgcgtcctttgtgaagttccattcagaagccagcgggtggtgatcatgctcgtcaaggcc 215
Db 491 GTCGAGGGCTGTGGCGCTGCACATCCCGAGGACCTGCGCGTCCGCTGGTTCGAGCGC 432
QY 216 ggcgcgcagtgtagccagaccatcgcgacgctcgacgtcacitggagcagggcgactgc 275
Db 431 GAACCGGAGTGTGTCATGCTGCCCGCCGACACCCCGCGCGCGCGCGCGCGCTCGTCCGC 372
QY 276 atcatcgatggggggaacgagtggt 300
Db 371 CTCGCCGACCTCGCGGACGACCGGT 347

```

RESULT 14

```

US-08-272-882D-1/c
; Sequence 1, Application US/08272882D
; Patent No. 5516685
; GENERAL INFORMATION:
; APPLICANT: Lichenstein, Henri
; TITLE OF INVENTION: Isolation and Characterization of No. 5516685e1
; TITLE OF INVENTION: Protease from Streptomyces Lividans
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/272,882D
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy
; REFERENCE/DOCKET NUMBER: A-199A
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2261 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-272-882D-1

```

```

Query Match      8.2%; Score 33.8; DB 1; Length 2261;
Best Local Similarity 47.8%; Pred. No. 2.1;
Matches 98; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

```

QY 96 gacgagaccgtgcagcgtgcgaagcgagaggaacaccttcccgctctacggttccatgac 155
Db 551 GCCAACGCGCTGTCGCGCGGGTGACCGAGGTCAGCTGACGCTGCGCTTCGTGCGACGAG 492
QY 156 cccgcgtcctttgtgaagttccattcagaagccagcgggtggtgatcatgctcgtcaagcc 215
Db 491 GTCGAGGGCTGTGCGCTGCACATCCCGAGGACCTGCGCGTCCGCGCTCGTGGTCCGCGC 432
QY 216 ggcgcgcagttgaccagaccatcgcgacgtctcgacgtcacttggagcagggcgactgc 275
Db 431 GAACCGGAGTGTGTCATGCTGCCCGCCGACACCCCGCGCGCGCGCGCTCGTCCGC 372
QY 276 atcatcgatggggggaacgagtggt 300
Db 371 CTCGCCGACCTCGCGGACGACCGGT 347

```

RESULT 15

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:43:19 ; Search time 320.8 Seconds
(without alignments)
268.163 Million cell updates/sec

Title: US-09-300-482-27
Perfect score: 229

Sequence: 1 cagacctatttttctgtc.....cttcaatttatggggaacaa 229

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
C 1	33.6	14.7	1105	19 296449
C 2	32.4	14.1	910715	20 X20248
C 3	30.4	13.3	1026	18 X20248
C 4	30.4	13.3	15249	18 V74413
C 5	29.8	13.0	4434	20 X61678
C 6	29.8	13.0	4485	20 X61677
C 7	29.8	13.0	116277	20 X20249
C 8	29.6	12.9	1435	17 T43709
C 9	29.6	12.9	2580	18 T70838
C 10	29.6	12.9	2589	18 T61590
C 11	29.6	12.9	3532	18 T72711
C 12	29.6	12.9	3532	20 Z22143
C 13	29.6	12.9	3732	19 V55040
C 14	29.4	12.8	388	20 V90010
C 15	29	12.7	72604	20 Z10752
C 16	28.8	12.6	618	21 A16054
C 17	28.8	12.6	625	20 V89678
C 18	28.4	12.4	360	18 V77318
C 19	28.4	12.4	1530	19 V27310
C 20	28.4	12.4	1579	17 T44594
C 21	28.4	12.4	1664976	19 V21209
C 22	28	12.2	750	20 X30326
C 23	28	12.2	2076	21 Z46833
C 24	28	12.2	2759	9 N80149
C 25	27.8	12.1	1969	20 X60804
C 26	27.6	12.1	995	19 X14581
C 27	27.4	12.0	964	18 T79676
C 28	27.4	12.0	1041	18 T79676
C 29	27.4	12.0	1050	18 T79674
C 30	27.4	12.0	1599	17 T05834
C 31	27.4	12.0	8010	17 T13413
C 32	27.4	12.0	10726	19 V52199
C 33	27.4	12.0	22080	20 X06751
C 34	27.2	11.9	416	20 V08862
C 35	27.2	11.9	612	20 X20494
C 36	27.2	11.9	695	19 Z27267
C 37	27.2	11.9	1070	11 Q05965
C 38	27.2	11.9	1705	19 Z96331
C 39	27.2	11.9	1705	19 V37382
C 40	27.2	11.9	3246	20 V08861
C 41	27.2	11.9	3247	20 X78784
C 42	27	11.8	251	16 T22960
C 43	27	11.8	756	15 Q73729
C 44	27	11.8	957	14 Q36645
C 45	27	11.8	1008	17 T16482

ALIGNMENTS

RESULT 1
296449/c
ID 296449 standard; DNA; 1105 BP.
XX AC 296449;
XX DT 10-APR-2000 (first entry)
XX DE S. pneumoniae derived DNA from ORF #277.
XX KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
XX KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX OS Streptococcus pneumoniae.
XX PN W09806734-A1.
XX PD 19-FEB-1998.
XX PF 15-AUG-1997; 97WO-US14436.
XX PR 16-AUG-1996; 96US-0024022.
XX PR (SMK) SMITHKLINE BEECHAM CORP.
XX PA Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
XX PI Stodola RK;
XX WPI; 1998-159452/14.
XX DR P-PSDB; Y86134.
XX PT Streptococcus pneumoniae proteins and related DNA - useful for
XX PT screening compounds for antibacterial activity
XX PS Claim 4; Page 291; 640pp; English.

Human HIAP-2 codin
EST clone C2326.
Genomic sequence o
Human colon cancer
EST clone DA505.
Staphylococcus aur
Arabidopsis leaf a
AF1q cDNA involved
Methanococcus jann
DNA encoding a hum
R. sphaeroides gen
Insert of plasmid
Human secreted pro
H. pylori GHP0 167
BRCA2 cancer susce
BRCA2 cancer susce
BRCA2 cancer susce
Recombinant papill
Human papillomavir
Streptococcus pneu
Salmonella enteric
Human cancer-relat
Human secreted pro
Arabidopsis thalia
Sequence encoding
S. pneumoniae deri
Streptococcus pneu
Human cancer-relat
Human CTIP cDNA.
Human gene signatu
Rape abscission/de
Soluble human inte
Soluble interleuki

XX This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see Z96173-Z96494) and their encoded proteins (see
 CC Y85792-Y86182). The DNA, vectors and host cells described in the method
 CC of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX Sequence 1105 BP; 303 A; 244 C; 193 G; 365 T; 0 other;

Query Match 14.7%; Score 33.6; DB 19; Length 1105;
 Best Local Similarity 52.1%; Pred. No. 0.17;
 Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 65 acaagaatagccttgctggtgctggttaattggtgcaaaatctggcactcaatattgct 124
 Db 284 ACTGGCTCGGCTTGTGTTAGCGGATTTTCCTGCTAAACACAGCTCCACAAACACCT 225
 Qy 125 tgaagaaggcttcccaattccggttaacacggaacatttccaaggttattgggccata 184
 Db 224 GTCRAAGAGATCGAATTTGACGCTGAAGTCGCAGCTGTTTCCCAAGGATTTGGTATCCGA 165
 Qy 185 agacgaagcaaacaggagagaaa 208
 Db 164 AAGGAGAGTGAACAGGAAGAAA 141

RESULT 2

X20248/c
 ID X20248 standard; DNA; 910715 BP.

XX AC X20248;

XX DT 04-MAY-1999 (first entry)

XX DE Borrelia burgdorferi polynucleotide sequence #1.

XX KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.

XX OS Borrelia burgdorferi.

XX PN W09858943-A1.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US12764.

XX PR 03-SEP-1997; 97US-0057483.

XX PR 20-JUN-1997; 97US-0050359.

XX PR 22-JUL-1997; 97US-0053344.

XX PR 22-JUL-1997; 97US-0053377.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PA (MEDI-) MEDIMUNE INC.

XX PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;

XX PI White OR;

XX DR WPI; 1999-081217/07.

XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 products for the detection, diagnosis, characterisation, prevention

PT and therapy of infections, particularly Lyme disease

XX Claim 1; Page 157-671; 1128pp; English.

XX X20248 to X20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (8b). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.

XX Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;

Query Match 14.1%; Score 32.4; DB 20; Length 910715;
 Best Local Similarity 49.4%; Pred. No. 4.4;
 Matches 84; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

Qy 57 aaccacacaaagaatagccttgctggtgctggttaattggtgcaaaatctggcactca 116
 Db 171375 AATTTAAACAAAGATAAATCTTGATGTTGCTGCTATTGTCGAAAAAATAAATG 171316
 Qy 117 atattgcttgaagggttcccaattccggttaacacggaacatttccaaggttatt 176
 Db 171315 ATATTGTTTATAGCATGCTTTTAAAGGTTGAAATGAAGATAAACCTGTTTCATTTTGATT 171256
 Qy 177 gggccataagacgaagcaaacaccaggaagaaacaccttccatttatggggaa 226
 Db 171255 GTGTTATCAATAAATTACATCTGAAATAAATCTTTCAAGTAATGAAAAA 171206

RESULT 3

Z51822
 ID Z51822 standard; DNA; 1026 BP.

XX AC Z51822;

XX DT 04-JUL-2000 (first entry)

XX DE Staphylococcus aureus glycoprotease (gcp) polynucleotide.

XX KW Staphylococcus aureus glycoprotease; gcp; antibody; vaccine formulation;
 KW Helicobacter pylori infection; gastrointestinal carcinoma; gastric ulcer;
 KW gastritis; osteomyelitis; septic arthritis; septic thrombophlebitis;
 KW acute bacterial endocarditis; Staphylococcus food poisoning;
 KW scalded skin syndrome; toxic shock syndrome; antibacterial; cytostatic;
 KW antiulcer; antiarthritic; vulnery; immunosuppressive; ss.

XX OS Staphylococcus aureus.

XX PH Key Location/Qualifiers

XX FT CDS 1..1026 a
 XX FT /tag= a
 XX FT /product= "Glycoprotease polypeptide"

XX PN W0200013694-A1.

XX PD 16-MAR-2000.

XX PF 23-AUG-1999; 99WO-US19273.

XX PR 08-SEP-1998; 98US-0149624.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Palmer LM, Kosmatka AL, Traini CM, Warren RL;

XX DR WPI; 2000-256855/22.

XX DR P-PSDB; Y70524.

Db 1298 agggactgttcgtacattctctctaaagaaaaatagctctatttttaacctgcataa 1357
 Qy 128 aaaggcttcccaattccggttaacacggaacacatttcccaagggtatttgggcataaga 187
 Db 1358 aaaggtctttaaattgttgaaacacttgaagccatctaaagtaaaagggaattatga 1417

RESULT 9

T70838
 ID T70838 standard; cDNA; 2580 BP.

XX AC T70838;

XX DT 02-SEP-1997 (first entry)

XX DE Human apoptosis inhibitor hiap-2 cDNA.

XX KW Apoptosis inhibitor; HIAP-2; HIV; AIDS; neurodegeneration;

XX KW myelodysplastic syndrome; ischaemia; myocardial infarction; stroke;

XX KW reperfusion injury; toxin-induced liver disease; gene therapy;

XX KW diagnosis; ds.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 238..2094

XX FT /*tag= a

XX PN W09706255-A2.

XX PD 20-FEB-1997.

XX PF 05-AUG-1996; 96WO-IB01022.

XX PR 22-DEC-1995; 95US-0576956.

XX PR 04-AUG-1995; 95US-0511485.

XX PA (UYOT-) UNIV OTTAWA.

XX PI Baird S, Korneluk RG, Liston P, Mackenzie AE;

XX DR WPI: 1997-154262/14.

XX DR P-PSDB; W19583.

XX PT Nucleic acid encoding an inhibitor of apoptosis polypeptide - used
 to inhibit apoptosis in e.g. HIV or AIDS patients, and for detection
 of susceptibility to apoptotic disease

XX PS Claim 12; Page 74-75; 219pp; English.

XX CC Human xiap, hiap-1 and hiap-2 genes, and murine xiap, hiap-1 and
 hiap-2 genes (T70836-41) respectively code for a new class of
 mammalian proteins (W19581-86) that are inhibitors of apoptosis
 (IAP). The hiap-1 and hiap-2 genes were cloned by screening a
 human liver library with a probe including the entire xiap coding
 region. They were assigned to chromosome 11q22-23 by FISH. IAP
 nucleic acids can be used to express IAP polypeptides in cells and
 animals to inhibit apoptosis, and as primers and probes to identify
 and isolate additional IAP genes, as well as in methods for treating
 diseases and disorders involving apoptosis (anti-apoptotic gene
 therapy).

XX SQ Sequence 2580 BP; 800 A; 445 C; 534 G; 799 T; 2 other;

Query Match 12.9%; Score 29.6; DB 18; Length 2580;

Best Local Similarity 47.8%; Pred. No. 4.5;

Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 8 tatttttctgctattgttcaaaatttcaggagagataattatgcgtcaaccacaaca 67

Db 2006 tagtatgccagggaatgtgccctctctctaaagaaatgcctatttgcagggtataatca 2065

Qy 68 agaataggccttgcgtgattgctgttaattggcgaatactcggcactcaattgtctga 127
 Db 2066 agggtagtctgttcgtacattctctctctaaagaaaaatgctctatttttaacctgcataa 2125
 Qy 128 aaagggtcttcccaattccggttaacacggaacacatttcccaagggtatttgggcataaga 187
 Db 2126 aaaggtctttaaattgttgaaacacttgaagccatctaaagtaaaagggaattatga 2185

RESULT 10

T61590

ID T61590 standard; cDNA; 2589 BP.

XX AC T61590;

XX DT 22-JUL-1997 (first entry)

XX DE Human c-IAP1.

XX KW IAP; inhibitor; apoptosis; RING finger domain; restinosis;

XX KW myocardial infarction; nephritis; HIV; ss.

XX OS Homo sapiens.

XX PN W09706182-A1.

XX PD 20-FEB-1997.

XX PF 06-AUG-1996; 96WO-US12860.

XX PR 08-DEC-1995; 95US-0569749.

XX PR 08-AUG-1995; 95US-0512946.

XX PA (TULA-) TULARIK INC.

XX PI Goeddel DV, Rothe M;

XX DR WPI: 1997-154209/14.

XX DR P-PSDB; W13545.

XX PT Nucleic acids encoding cellular inhibitor of apoptosis proteins -

useful for apoptosis regulation in cells to reduce or increase

apoptosis and for pharmacological screening

XX PS Claim 1; Page 16-18; 35pp; English.

XX CC The human cellular inhibitor of apoptosis proteins (c-IAP1/2 -
 T61590/T61591) comprise a series of defined structural domain
 repeats and/or a RING finger domain; in particular, at least two of
 a first domain repeat (W13547 or W13548), a second domain repeat
 (W13549 or W13550), and a third domain repeat (W13551 or W13552)
 CC and/or a RING finger domain (W13553 or W13554), or a consensus
 sequences derived from these human genes.

XX CC The nucleic acid is used for recombinant prodn. of human cellular

inhibitor of apoptosis protein which modulates apoptosis

regulation. The nucleic acids are useful in therapies where

increased cell-specific apoptosis is desired, e.g. in restinosis,

inflammatory disease states, myocardial infarction, glomerular

nephritis, transplant rejection and infectious diseases, e.g. HIV.

XX They can also be used in conditions requiring a reduction in

apoptosis.

XX SQ Sequence 2589 BP; 844 A; 436 C; 522 G; 787 T; 0 other;

Query Match 12.9%; Score 29.6; DB 18; Length 2589;

Best Local Similarity 47.8%; Pred. No. 4.5;

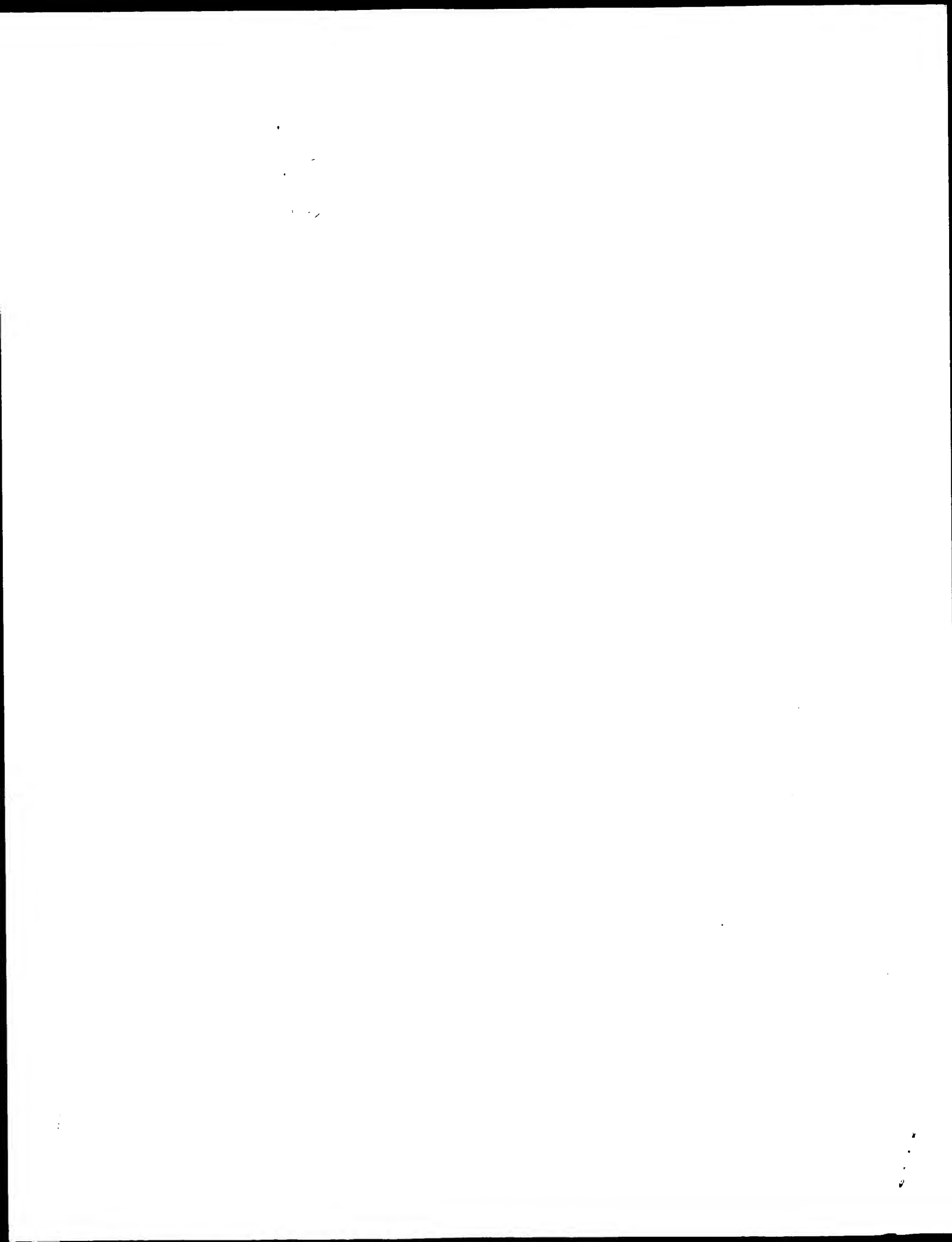
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 8 tatttttctgctattgttcaaaatttcaggagagataattatgcgtcaaccacaaca 67

Db 1963 tagtatgccagggaatgtgccctctctctaaagaaatgcctatttgcagggtataatca 2022

8 tatttttctgcatcttgctcaaatccaggagattaattatgcgtcaccacaca 67
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2928 tagtatgccgaagaatgycacctctctaagaaatgcctatttcaggggataatca 2987

8 tatttttttgtoattgctccaatttcaggagattaattatgctgaaccccaaca 67
228 tagtatgccaggaaatgccccctctctaagaaaatgcctatttgcagggtataatca 2987
68 agaataggcccttgctggattggctgttaattgggcaaaaatctggcactcaattgcttga 127
2988 aggtactgtctgcataattctctcttaagaagaaatagtcctataatttcaacctcataa 3047



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:33:31 ; Search time 189.35 Seconds
(without alignments)
182.912 Million cell updates/sec

Title: US-09-300-482-27
Perfect score: 229
Sequence: 1 cagaccttatttttctgtc.....cttcaatttatgggaacaa 229

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues
Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29.6	12.9	1435	6	PCT-US95-05922A-1
2	29.6	12.9	2580	4	US-08-511-485-7
3	29.6	12.9	2589	6	PCT-US96-12860-1
4	29.6	12.9	3532	4	US-09-205-204-1
5	29.6	12.9	3732	5	US-09-212-971-7
6	28.4	12.4	270	2	US-08-589-080-3
7	28.4	12.4	1629	2	US-08-589-080-2
8	27.4	12.0	964	5	US-08-755-587-23
9	27.4	12.0	1041	5	US-08-755-587-22
10	27.4	12.0	1050	5	US-08-755-587-21
11	27.4	12.0	1599	5	US-08-737-336-5
12	27.4	12.0	8010	6	PCT-US95-11859-2
13	27	11.8	756	3	US-08-530-165-1
14	27	11.8	1008	1	US-07-904-073-1
15	27	11.8	1008	1	US-07-904-071-1
16	27	11.8	1770	3	US-08-381-603-1
17	27	11.8	1770	6	PCT-US94-02414-1
18	27	11.8	1770	6	PCT-US96-08899-1
19	27	11.8	3011	1	US-07-821-716-1
20	26.6	11.6	714	5	US-08-643-704A-44
21	26.6	11.6	816	5	US-08-643-704A-47
22	26.6	11.6	900	5	US-08-643-704A-46
23	26.6	11.6	1236	5	US-08-643-704A-48
24	26.6	11.6	4649	7	5183745-1
25	26.6	11.6	6443	7	5183745-5
26	26.6	11.6	9432	1	US-08-277-231A-1

C 27	26.6	11.6	9432	3	US-08-473-750-4	Sequence 4, Appli
C 28	26.6	11.6	9432	4	US-08-477-326-4	Sequence 4, Appli
C 29	26.6	11.6	17041	1	US-08-076-011-1	Sequence 1, Appli
C 30	26.2	11.4	1194	5	US-08-929-329-3	Sequence 3, Appli
C 31	26	11.4	2095	2	US-08-405-230-8	Sequence 8, Appli
C 32	26	11.4	2095	3	US-08-910-990-8	Sequence 8, Appli
C 33	26	11.4	2564	5	US-08-276-968A-19	Sequence 19, Appli
C 34	25.8	11.3	6973	2	US-08-478-370-1	Sequence 1, Appli
C 35	25.6	11.2	816	5	US-08-968-563-8	Sequence 8, Appli
C 36	25.4	11.1	722	2	US-08-698-551-5	Sequence 5, Appli
C 37	25.4	11.1	722	3	US-08-602-228-5	Sequence 5, Appli
C 38	25.4	11.1	722	3	US-08-649-341A-5	Sequence 5, Appli
C 39	25.4	11.1	722	3	US-08-494-440B-5	Sequence 5, Appli
C 40	25.4	11.1	722	3	US-08-533-901B-5	Sequence 5, Appli
C 41	25.4	11.1	722	3	US-08-839-032A-5	Sequence 5, Appli
C 42	25.4	11.1	722	4	US-08-839-031A-5	Sequence 5, Appli
C 43	25.4	11.1	722	6	PCT-US95-12724-5	Sequence 5, Appli
C 44	25.4	11.1	1018	1	US-08-181-271A-30	Sequence 30, Appli
C 45	25.4	11.1	1018	1	US-08-449-315-30	Sequence 30, Appli

ALIGNMENTS

RESULT 1
PCT-US95-05922A-1
; Sequence 1, Application PC/TUS9505922A
; GENERAL INFORMATION:
; APPLICANT: HE, ET AL.
; TITLE OF INVENTION: Human Inhibitor of Apoptosis Gene 1
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05922A
; FILING DATE: 11 MAY 1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-292
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 BASE PAIRS
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: CDNA
PCT-US95-05922A-1

Query Match 12.9%; Score 29.6; DB 6; Length 1435;
Best Local Similarity 47.8%; Pred. No. 0.64;
Matches 86; Conservative 94; Indels 0; Gaps 0;
QY 8 tatttttgcattgtctcaatttcaggagattaattatgcgtcaccccaaca 67

NUMBER OF SEQ ID NOS: 47
SEQ ID NO 1
LENGTH: 3532
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1160)..(3016)
US-09-205-204-1

Query Match 12.9%; Score 29.6; DB 4; Length 3532;
Best Local Similarity 47.8%; Pred. No. 0.97;
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 8 tatttttctgcttgccttcaaaatttcaggagatttaattatgctcaacccacaaca 67
Db 2928 tagtatgcaggaaatgtgcccttctcctaagaaatgccctatttgcagggtataatca 2987
QY 68 agaataggccttgcgtgattggctgtaattggcgaataatctggcactcaaatattgcttga 127
Db 2988 agggtaactgttcgtacattctctcttaagaaaaatagtctctattttaacctgcataa 3047
QY 128 aaagggtcccaattccggttaacaacgcaaccatttcccaaggttatttgggccataaga 187
Db 3048 aaaggtctttaaataattgttgaaacacttgaagccatctaagtaaaaaagggaattatga 3107

RESULT 5
US-09-212-971-7
Sequence 7, Application US/09212971B
Patent No. 6107041
GENERAL INFORMATION:
APPLICANT: Korneluk, Robert G
APPLICANT: MacKenzie, Alexander E
APPLICANT: Liston, Peter
APPLICANT: Baird, Stephen
APPLICANT: Tsang, Benjamin K
APPLICANT: Pratt, Christine
TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
TITLE OF INVENTION: DISEASE
FILE REFERENCE: 07891/009002
CURRENT APPLICATION NUMBER: US/09/212.971B
CURRENT FILING DATE: 1998-12-16
EARLIER APPLICATION NUMBER: 60/017,354
EARLIER FILING DATE: 1996-04-26
EARLIER APPLICATION NUMBER: 60/030,590
EARLIER FILING DATE: 1996-11-14
EARLIER APPLICATION NUMBER: 08/800,929
EARLIER FILING DATE: 1997-02-13
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 3732
TYPE: DNA
ORGANISM: Homo sapiens
US-09-212-971-7

Query Match 12.9%; Score 29.6; DB 5; Length 3732;
Best Local Similarity 47.8%; Pred. No. 0.99;
Matches 86; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 8 tatttttctgcttgccttcaaaatttcaggagatttaattatgctcaacccacaaca 67
Db 3148 tagtatgcaggaaatgtgcccttctcctaagaaatgccctatttgcagggtataatca 3207
QY 68 agaataggccttgcgtgattggctgtaattggcgaataatctggcactcaaatattgcttga 127
Db 3208 agggtaactgttcgtacattctctcttaagaaaaatagtctctattttaacctgcataa 3267
QY 128 aaagggtcccaattccggttaacaacgcaaccatttcccaaggttatttgggccataaga 187

Db 3268 aaaggtctttaaataattgttgaaacacttgaagccatctaagtaaaaaagggaattatga 3327

RESULT 6

US-08-589-080-3
Sequence 3, Application US/08589080
Patent No. 5808026
GENERAL INFORMATION:
APPLICANT: COHEN, Amos
APPLICANT: TSE, William
APPLICANT: ZHU, Weimin
TITLE OF INVENTION: SEQUENCE OF Aflq CDNA
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/589,080
FILING DATE: 23-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9501386.8
FILING DATE: 23-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rea, Teresa Stanek
REGISTRATION NUMBER: 30,427
REFERENCE/DOCKET NUMBER: 024916-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-589-080-3

Query Match 12.4%; Score 28.4; DB 2; Length 270;
Best Local Similarity 62.9%; Pred. No. 0.76;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 146 ggtaaacaaacggaaccatttccaaaggttatttggccataagcaacacccagggaagg 205
Db 117 GGTCAAAGACAGCAGCGTGTGGCAAAATGATCGGGCAAGCAACTGCAGCAGCAGGAGAA 176

QY 206 aaaccttcaa 215

Db 177 AAACCTGAA 186

RESULT 7

US-08-589-080-2
Sequence 2, Application US/08589080
Patent No. 5808026
GENERAL INFORMATION:
APPLICANT: COHEN, Amos
APPLICANT: TSE, William
APPLICANT: ZHU, Weimin
TITLE OF INVENTION: SEQUENCE OF Aflq CDNA
NUMBER OF SEQUENCES: 6


```

; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 503..541
; FEATURE:
; NAME/KEY: exon
; LOCATION: 501..541
US-08-755-587-22

```

```

Query Match      12.0%; Score 27.4; DB 5; Length 1041;
Best Local Similarity 53.2%; Pred. No. 3.1;
Matches 58; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 2 agacctatttttctgtcatttgccttcaaaatttcaggagatttaattatgcgtcaacc 61
   |||||
Db 742 ATAGTTTATTTTATAGAAATGCCCTGATCAATTAAGGATATATTTTTCATTCACTCT 683
   |||||

Qy 62 acaacaagaataggccttgcgttgattgcttgaatgggcaaaatctgg 110
   |||||
Db 682 TTATTTCACTTAACGTTTACTGTAATTCCTGCTGATGAGGAGAGATGCTAG 634
   |||||

```

```

RESULT 10
US-08-755-587-21/c
; Sequence 21, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0

```

```

; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1050 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 502...550
; FEATURE:
; NAME/KEY: exon
; LOCATION: 501..550
US-08-755-587-21

```

```

Query Match      12.0%; Score 27.4; DB 5; Length 1050;
Best Local Similarity 53.2%; Pred. No. 3.1;
Matches 58; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 2 agacctatttttctgtcatttgccttcaaaatttcaggagatttaattatgcgtcaacc 61
   |||||
Db 883 ATAGTTTATTTTATAGAAATGCCCTGATCAATTAAGGATATATTTTTCATTCACTCT 824
   |||||

Qy 62 acaacaagaataggccttgcgttgattgcttgaatgggcaaaatctgg 110
   |||||
Db 823 TTATTTCACTTAACGTTTACTGTAATTCCTGCTGATGAGGAGAGATGCTAG 775
   |||||

```

```

RESULT 11
US-08-737-336-5
; Sequence 5, Application US/08737336
; Patent No. 6013262
; GENERAL INFORMATION:
; APPLICANT: FRAZER, Ian
; APPLICANT: ZHOU, Jian
; TITLE OF INVENTION: RECOMBINANT PAPILLOMA VIRUS L1
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,336
; FILING DATE: 16-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/AU95/00292
; FILING DATE: 17-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM 5667
; FILING DATE: 17-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 65064/118
; TELECOMMUNICATION INFORMATION:

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-6734
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US95-11859-2

```


; LOCATION: 1..1008
US-07-904-071-1

Query Match 11.8%; Score 27; DB 1; Length 1008;
Best Local Similarity 50.4%; Pred. No. 4.2;
Matches 66; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy	51	cacacacagaataggcccttgctggattggctgttaatgggcaaaaatctggcactcaatat	120
Db	226	CAACACAAAGAGAAACTTTGGTTTCTGCTAAGGTGGAGGATTCAGGACATTACTAT	285
Qy	121	tgctgaaaagggttccccaattccggttaacaacggaaccatttccaaggttattgggc	180
Db	286	TGCCTGGTAGAATTCTCTTACTGCTCAGAAATTAATAAGTGCRAAATTTGTGGAG	345
Qy	181	cataagacgaa	191
Db	346	AATGAGCCTAA	356

Search completed: November 4, 2000, 13:33:34
Job time: 16650 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:52:05 ; Search time 4352.3 Seconds
(without alignments)
325.314 Million cell updates/sec

Title: US-09-300-482-27

Perfect score: 229

Sequence: 1 cagacottatttttctgtc.....cttcaatttatggggaacaa 229

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estom:*
68: em_estov1:*
69: em_estov2:*
70: em_estpl1:*
71: em_estpl2:*
72: em_estpl3:*
73: em_estpl4:*
74: em_estpl5:*
75: em_estro1:*
76: em_estro2:*
77: em_estro3:*
78: em_estro4:*
79: em_estro5:*
80: em_estro6:*
81: em_estro7:*
82: em_estro8:*
83: em_estro9:*
84: em_estro10:*
85: em_estro11:*
86: em_estro12:*
87: em_estro13:*
88: gb_gss1:*
89: gb_gss2:*
90: gb_gss3:*
91: gb_gss4:*
92: em_gss1:*
93: em_gss2:*
94: em_gss3:*
95: em_gss4:*
96: gb_gss5:*
97: gb_gss6:*
98: gb_gss7:*
99: gb_gss8:*
100: gb_gss9:*
101: em_gss5:*
102: em_gss6:*
103: em_gss7:*
104: em_gss8:*
105: em_gss9:*
106: em_gss10:*
107: em_gss11:*
108: gb_gss10:*
109: gb_gss11:*
110: em_gss12:*
111: gb_gss12:*
112: gb_gss13:*
113: gb_gss14:*
114: gb_gss15:*
115: gb_gss16:*
116: gb_gss17:*

117: gb_gss18:*
 118: gb_gss19:*
 119: em_gss13:*
 120: gb_gss20:*
 121: gb_gss21:*
 122: gb_gss22:*
 123: gb_gss23:*
 124: gb_gss24:*
 125: em_gss14:*
 126: em_gss15:*
 127: em_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	139.4	60.9	561	32	BE022117	BE022117 sm67f12.y
2	113.4	49.5	532	24	AW719787	AW719787 LJNEST10A
3	110.4	48.2	430	18	AV406682	AV406682 AV406682
4	108.6	47.4	653	24	AW695821	AW695821 NF099E09S
5	101.4	44.1	499	32	BE020243	BE020243 sm42d04.y
6	96.2	42.0	573	23	AW586409	AW586409 EST318032
7	96.4	41.9	430	14	AL368071	AL368071 MCB22A07
8	89.6	39.1	420	14	AL362847	AL362847 MCB212H03
9	87.4	38.2	337	14	AL367204	AL367204 MCB212H03
10	87.4	38.2	479	14	AL365745	AL365745 MCB212H03
11	87.4	38.2	481	14	AL365743	AL365743 MCB212H03
12	87.4	38.2	630	23	AW688619	AW688619 NF009F08S
13	85.6	37.4	321	18	AV410206	AV410206 AV410206
14	85.6	37.4	379	18	AV410953	AV410953 AV410953
15	85.6	37.4	394	18	AV425788	AV425788 AV425788
16	82.2	35.9	606	21	AW299197	AW299197 EST306007
17	80.6	35.2	548	14	AL368600	AL368600 MCB25E04
18	79.4	34.7	733	34	A1055038	A1055038 coau0002N
19	78.4	34.1	332	34	BE204727	BE204727 EST397403
20	78.34	34.1	504	35	BE450814	BE450814 EST401701
21	74.6	32.6	361	24	AW760460	AW760460 s150f03.y
22	74.2	32.4	462	39	T04690	T04690 737 Lambda-
23	74.2	32.4	530	19	AV526953	AV526953 AV526953
24	72.6	31.7	528	19	AV550009	AV550009 AV550009
25	72.6	31.7	528	19	AV550009	AV550009 AV550009
26	72.2	31.5	434	21	AW255063	AW255063 ML1358 pe
27	71.6	31.3	375	19	AV540033	AV540033 AV540033
28	70.8	30.9	498	35	BE341175	BE341175 EST345276
29	67.7	29.3	638	21	AW254720	AW254720 ML1006 pe
30	66.6	28.8	158	35	BE343110	BE343110 EST395954
31	64.6	28.2	592	23	AW560451	AW560451 EST315499
32	63.4	27.7	382	35	BE403238	BE403238 WHE0430.A
33	63.4	27.7	382	35	BE406322	BE406322 WHE0413.e
34	63.4	27.7	401	35	BE428880	BE428880 MTD011.H0
35	63.4	27.7	460	23	AW677090	AW677090 DGL 4.B11
36	63.4	27.7	461	23	AW286858	AW286858 LGL 222.D
37	63.4	27.7	462	21	AW286839	AW286839 LGL 222.B
38	63.4	27.7	463	21	AW286879	AW286879 LGL 222.F
39	63.4	27.7	472	36	BE497862	BE497862 WHE0957.D
40	63.4	27.7	516	35	BE363025	BE363025 DGL 9.A10
41	63.4	27.7	519	23	AW563648	AW563648 LGL 245.A
42	63.4	27.7	530	35	BE363077	BE363077 DGL 9.F06
43	61.8	27.0	354	36	BE517810	BE517810 WHE0803.E
44	61.8	27.0	454	36	BE495944	BE495944 WHE1259.B
45	61.8	27.0	474	36	BE497880	BE497880 WHE0957.B

ALIGNMENTS

RESULT 1
 BE022117

LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BE022117 561 bp mRNA
 sm67f12.y1 Gm-c1028 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-c1028-9000 5' similar to TR:022111 022111 6-PHOSPHOGLUCONATE
 DEHYDROGENASE ; mRNA sequence.
 BE022117
 BE022117.1 GI:8284549
 EST.
 soybean.

REFERENCE
 AUTHORS

Glycine max
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 561)

TITLE
 JOURNAL
 COMMENT

Shoemaker, R., Keim, P., Vodkin, L., Erpelbing, J., Coryell, V., Khanna
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.
 Public Soybean EST Project
 Unpublished (1999)
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 398.
 Location/Qualifiers
 1..561
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1028-9000"
 /clone_lib="Gm-c1028"
 /tissue_type="roots of 'Supernod' plants"
 /lab_host="DH10B"
 /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
 XhoI; The mRNA was isolated from roots of Glycine max
 'Supernod' plants generously donated by Dr. Gary Stacey.
 The seedlings were inoculated with Bradyrhizobium
 japonicus, strain USDA110 prior to harvest. Stratagene's
 cDNA synthesis kit (catalog number 200401) was used to
 synthesize the cDNA. First-strand synthesis was performed
 with 5-methyl dCTP, hence the ligated cDNA was
 hemimethylated. A modification of Stratagene's
 first-strand synthesis primer was used. An 'anchor'
 nucleotide (V-A-C, or G) was added to the 3' end of the
 primer (GAGAGAGAGAGAGAGAGAGACTGCTCGAG(T)18V) to anchor
 the primer at the 5' end of the poly(A) tract. After
 second-strand synthesis, the cDNA ends were filled in with
 cloned Pfu DNA polymerase, ligated to EcoRI adapters and
 subsequently phosphorylated. The XhoI site within the
 first-strand synthesis primer was then restricted by
 digestion with XhoI; all XhoI sites in the cDNA would be
 protected by their hemimethylated status. The cDNA
 constructs were size-fractionated with a 500bp cutoff,
 using GibcoBRL life Technologies' cDNA size fractionation
 column. The column eluent was then ligated into
 Stratagene's pBluescript II XR Predigested vector
 (pBluescript II SK(+)) that has been digested with EcoRI
 and XhoI, and phosphorylated by Stratagene. Both the
 white and blue colonies appear to contain recombinant
 plasmids with cDNA inserts, based on size (n=25). This
 library was constructed by Dr. Paul Keim and Dr. Virginia
 Coryell."

FEATURES
 source

BASE COUNT 158 a 115 c 130 g 158 t
 ORIGIN

RESULT
AW695821

Qy 53 gctcaaccccaacaagaataggcccttgctggtattgctggttaatggggcaaaatctggca 112

```

Db 14 GCTCAACCCCTCAACAAGATACGCCCTTGCTGGACTGCTTATGCGCCCAAAATCTAGCA 73
Qy 113 ctcaatttctgtaaaagggttcccaattccggttaacaacgggaaccatttccaagt 172
Db 74 CTCAATATTGC-TGAGAAAGGCTTTCCCAATTTCTGTTTATAACCGCAACCACTTACAAGT 132
Qy 173 tattggccataagaagaacacaggaagaaacaccccttcaatttatgggaacaa 229
Db 133 TGATGAGACTGTGGACGACGACCAACATGAAGGAATCTCCAGTTATGGCTACCA 189

RESULT 6
AW586409 573 bp mRNA EST 20-MAR-2000
LOCUS EST3180332 MHAM Medicago truncatula/Glomus versiforme mixed EST
DEFINITION library cDNA clone pHAM-5403, mRNA sequence.
ACCESSION AW586409
VERSION AW586409.1 GI:7265923
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 573)
AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after colonization with
Glomus versiforme
JOURNAL Unpublished (2000)
COMMENT Contact: Marla J. Harrison
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name:N253596
TIGR sequence name:MTDAQ86TK
More information is available at...
'http://chrystle.tamu.edu/medicago'
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
FEATURES
source
Location/Qualifiers
1..573
/organism="Medicago truncatula/Glomus versiforme mixed EST
library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone_lib="pHAM-5403"
/clone="pHAM"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The cDNA was
directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in SOLR cells."
BASE COUNT 164 a 115 c 139 g 154 t 1 others
ORIGIN
Query Match 42.0%; Score 96.2; DB 23; Length 573;
Best Local Similarity 74.6%; Pred. No. 1.8e-19;
Matches 147; Conservative 0; Mismatches 48; Indels 2; Gaps 2;

Qy 21 atttgcataatttcaggagataattatgcgtccaccacacagaataagcgttg 80
Db 7 AATCTCTTCAATTTTCAGGAATAATATG-GCTCAACCACTTACAAGATAGCGCTTG 65

```

```

Qy 81 ctggattgcttaattgggcaaaatctggcactcaattgcttgaaagggtctcca 140
Db 66 CGGACTGGCTGTTATGCGCCAAATCTCGCACTCAATATTGC-CGACAAAGGATTCCCA 124
Qy 141 attccggttaacaacgggaaccatttccagggttattggccataagaacgaacacag 200
Db 125 ATTTCTGTTTATAACAGAACACATCAAAAGTTGACCAAACTGTGGAAACGACAAACAG 184
Qy 201 gaaggaacaccttcaatt 217
Db 185 GAGGGAATCTTCCACT 201

RESULT 7
AL368071 430 bp mRNA EST 03-AUG-2000
LOCUS MTBA22A07F1 MtBA Medicago truncatula cDNA clone MtBA22A07 T3, mRNA
DEFINITION sequence.
ACCESSION AL368071
VERSION AL368071.1 GI:9667824
SOURCE EST.
ORGANISM barrel medic.
Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 430)
AUTHORS Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O.,
Niebel,A., Carreau,Y., Chatagnier,O., Kahn,D., Glapinazzi-Pearson
,V. and Gamas,P.
Medicago truncatula ESTs from nitrogen-starved roots
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Moleculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
Mt-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Mtruncatula.html).
FEATURES
source
Location/Qualifiers
1..430
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MtBA22A07"
/clone_lib="MtBA"
/tissue_type="root tips"
/dev_stage="harvested after 3 days of N-starvation"
/note="Vector: pBluescript pSK; Site_1: EcoRI; Site_2:
XhoI; plants were grown in an aeroponic chamber for 14
days on nitrogen-rich medium followed by 3 days on N-free
medium. RNA was extracted from root tips (1-3 cm). cDNA
was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zapXR vector from
Stratagene and packaged using Gigapack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using ExAssist helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."
BASE COUNT 131 a 85 c 103 g 111 t
ORIGIN
Query Match 41.9%; Score 96; DB 14; Length 430;
Best Local Similarity 75.5%; Pred. No. 2e-19;
Matches 145; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

Qy 26 ctccaatttcaggagataattatgcgtccaccacacagaataagcgttgctga 85

```

Db	2	CTTCAATTTTCAGAAATACATTATG-GCTCAACCACTTACAGAAATAGGCTTGC CGGA	60
Qy	86	ttggtcttaattggcgaaatctggcactcaatattgcttgaagaaggtctcccaattcc	145
Db	61	CTGGCTGTTATGGGCCAAATTCGCACTCAATATTGC-CGACAAAGGATTTCCCAATTC	119
Qy	146	ggttaacaacggaaccatttcccaagttattggccataagaacgaacgaacgaagaag	205
Db	120	TGTTTATTAACAGAACACATCAAAAGTTGACGAACCTGTGTGACGAGCAAAACAGAGG	179
Qy	206	aaacctttcaatt	217
Db	180	AAATCTTCCACT	191
RESULT	8		
AL382847			
LOCUS	AL382847	420 bp	mRNA
DEFINITION	MBE10C10F1 MtBC Medicago truncatula cDNA clone M1BC10C10 T3, mRNA	EST	03-AUG-2000
ACCESSION	AL382847		
VERSION	AL382847		
KEYWORDS	EST.		
SOURCE	AL382847.1	GI:9682598	
ORGANISM	barrel medic.		
	Medicago truncatula		
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Medicago.		
REFERENCE	1 (bases 1 to 420)		
AUTHORS	Journet,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.		
TITLE	Medicago truncatula ESTs from endomycorrhizal roots		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Genoscope		
	Genoscope - Centre National de Sequencage		
	BP 191 91006 Evry cedex - France		
	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr		
	Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).		

BASE COUNT	128 a	83 c	102 g	106 t	1 others
ORIGIN	lungal origin.				

Query Match	39.1%	Score 89.6	DB 14	Length 420
Best Local Similarity	74.6%	Pred. No. 1.9e-17		
Matches 138	Conservative	0	Mismatches 45	Indels 2
Gaps				
QY	33	tttcaggagattaattatgcctcaaccccaacaaagaatagccttgctggattgctg	92	
Db	1	TTTCAGGAATACATTATG-GCTCAACCACTTACAAAGATAGGCCTTCGCCGACTGGCTG	59	
QY	93	ttatggggcaaaaactctggcactcaatattgcttgaaaggcttcccgaattcccggttaac	152	
Db	60	TTATGGGCCAAAATCTCGACTCAATATGC-CACNAAGGATTCCCAATTTCTGTTTAT	118	
QY	153	aacggaaccatttccaaggttattggggccataagacgaagcaaacagggaaggaaacctt	212	
Db	119	AACAGAACACATCAAAAGGTTGACGAAACTGTGGAACGAGCAAAACAGGAGGAATCTT	178	
QY	213	caatt 217		
Db	179	CCACT 183		

RESULT	9
AL3367204	
LOCUS	AL3367204 337 bp mRNA EST 03-AUG-2000
DEFINITION	MCBA12H03F1 MTBA Medicago truncatula cDNA clone MCB12H03 T3, mRNA sequence.
ACCESSION	AL3367204
VERSION	AL3367204
KEYWORDS	AL3367204.1 GI:9666957
SOURCE	EST.
ORGANISM	barrel medic. Medicago truncatula Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE	1 (bases 1 to 337)
AUTHORS	Journel,E.P., Crespeau,H., van-Tuinen,D., Gouzy,J., Jaillon,O., Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson,V. and Gamas,P.
TITLE	Medicago truncatula ESTs from nitrogen-starved roots
JOURNAL	Unpublished (2000)
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Contact : Pascal Gamas and Etienne-Pascal Journet. Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).

BASE COUNT	108 a	71 c	72 g	86 t
sequencing (genoscope, evry)				

ORIGIN

Query Match 38.2%; Score 87.4; DB 14; Length 337;
 Best Local Similarity 74.5%; Pred. No. 8.6e-17;
 Matches 123; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

QY 53 gctcaaccacacaaagaatagccttgctggattgctgtaataaggcacaatactggca 112
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 16 GCTCAACCACTTACAAGATAGCCCTGCCGACTGCTGTTATGCGCCAAATCTCGCA 75
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 113 ctcaatattgctgaaaggctcccaattccggttaacacgaaccatttccaaggt 172
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 76 CTCAATATTGC-CGACAAAGGATTCCCAATTTCTGTTTATAACAGAACATCAAGGT 134
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 173 tattggccataagaacgaacacgaagaagaaacattcaatt 217
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 135 TGACGAACTGTGGAACGACGAAACAGGAGGAGGAAATCTTCCACT 179
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10

AL365745 479 bp mRNA EST 03-AUG-2000
 LOCUS MTBA02B10F3 McBA Medicago truncatula cDNA clone MTBA02B10 T3, mRNA
 DEFINITION sequence.
 ACCESSION AL365745
 VERSION AL365745.1 GI:9665498
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE

1 (bases 1 to 479)
 JOURNAL JOURNAL

COMMENT

Medicago truncatula ESTs from nitrogen-starved roots

Unpublished (2000)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de

Biologie Moleculaire des Relations Plantes-Microorganismes,

CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :

Mt-est@toulouse.inra.fr Website :

http://sequence.toulouse.inra.fr/Mtruncatula.html).

Location/Qualifiers

1. .479

/organism="Medicago truncatula"

/cultivar="Jemalong"

/db_xref="taxon:3880"

/clone="MtBA02B10"

/clone_lib="McBA"

/tissue_type="root tips"

/dev_stage="harvested after 3 days of N-starvation"

/note="Vector: pBluescript PSK; Site.1: EcoRI; Site.2:

XhoI; Plants were grown in an aeroponic chamber for 14

days on nitrogen-rich medium followed by 3 days on N-free

medium. RNA was extracted from root tips (1-3 cm). cDNA

was prepared from polyA+ enriched RNA. The cDNA was

directionally ligated into Uni-zapR vector from

Stratagene and packaged using Gigapack Gold packaging

extracts. Plasmids containing cDNA inserts were

mass-excised from phage stocks using Exassit helper phage

and propagated in SOLR cells. Clone ordering and

sequencing was performed by the Centre National de

Sequencage (Genoscope, Evry, France)."

141 a 95 c 120 g 123 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 74.5%; Score 87.4; DB 14; Length 479;
 Matches 123; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

QY 53 gctcaaccacacaaagaatagccttgctggattgctgtaataaggcacaatactggca 112
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 13 GCTCAACCACTTACAAGATAGCCCTGCCGACTGCTGTTATGCGCCAAATCTCGCA 72
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 113 ctcaatattgctgaaaggctcccaattccggttaacacgaaccatttccaaggt 172
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 73 CTCAATATTGC-CGACAAAGGATTCCCAATTTCTGTTTATAACAGAACATCAAGGT 131
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 173 tattggccataagaacgaacacgaagaagaaacattcaatt 217
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 132 TGACGAACTGTGGAACGACGAAACAGGAGGAGGAAATCTTCCACT 176
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11

AL365743 481 bp mRNA EST 03-AUG-2000
 LOCUS MTBA02B09F3 MtBA Medicago truncatula cDNA clone MTBA02B09 T3, mRNA
 DEFINITION sequence.
 ACCESSION AL365743
 VERSION AL365743.1 GI:9665496
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;

Fabales; Fabaceae; Papilionoideae; Medicago.

1 (bases 1 to 481)

JOURNAL JOURNAL

Medicago truncatula ESTs from nitrogen-starved roots

Unpublished (2000)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de

Biologie Moleculaire des Relations Plantes-Microorganismes,

CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :

Mt-est@toulouse.inra.fr Website :

http://sequence.toulouse.inra.fr/Mtruncatula.html).

Location/Qualifiers

1. .481

/organism="Medicago truncatula"

/cultivar="Jemalong"

/db_xref="taxon:3880"

/clone="MtBA02B09"

/clone_lib="MtBA"

/tissue_type="root tips"

/dev_stage="harvested after 3 days of N-starvation"

/note="Vector: pBluescript PSK; Site.1: EcoRI; Site.2:

XhoI; Plants were grown in an aeroponic chamber for 14

days on nitrogen-rich medium followed by 3 days on N-free

medium. RNA was extracted from root tips (1-3 cm). cDNA

was prepared from polyA+ enriched RNA. The cDNA was

directionally ligated into Uni-zapR vector from

Stratagene and packaged using Gigapack Gold packaging

extracts. Plasmids containing cDNA inserts were

mass-excised from phage stocks using Exassit helper phage

and propagated in SOLR cells. Clone ordering and

sequencing was performed by the Centre National de

Sequencage (Genoscope, Evry, France)."

142 a 95 c 121 g 123 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 74.5%; Score 87.4; DB 14; Length 481;
 Matches 123; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

QY 53 gctcaaccacacagaagtagcttgctgattgctgtaattgagcgaattctggca 112
 Db 13 GCTCAACACACTTACAAGAATAAGGCTTGGCGGACTGGCTGTATGGGCCAAATCTCGCA 72

QY 113 ctcaatatgtctgaaaggcttcccaattcccggttaacacaggaaccatttccaaggt 172
 Db 73 CTCATATATGCG-CAACAAGGATCCCAATTCCTGTATATACAGACACATCAAGGT 131

QY 173 tattgggccaataagcagcaacacaggaaggaaaccttcaatt 217
 Db 132 TGACGAAACTGTGGAACGACGACAAACAGGAGGAAATCTTCCACT 176

RESULT 12
 AW688619 630 bp mRNA EST 15-JUN-2000
 LOCUS
 DEFINITION NF009F08ST1F1000 Developing stem Medicago truncatula cDNA clone
 NF009F08ST 5', mRNA sequence.
 ACCESSION AW688619
 VERSION AW688619
 KEYWORDS
 SOURCE barrel medic.
 ORGANISM
 Medicago truncatula
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE
 AUTHORS He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
 ,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
 ,R.A.
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 Medicago truncatula stem library
 JOURNAL Unpublished (2000)
 COMMENT Contact: Dixon RA
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA
 Tel: 580 221 7302
 Fax: 580 221 7380
 Email: radixon@noble.org
 Insert Length: 630 Std Error: 0.00
 Plate: 009 Row: F Column: 08
 Seq primer: TCACACAGGAACACAGCTATGAC.
 Location/Qualifiers
 1..630
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF009F08ST"
 /clone_lib="Developing stem"
 /tissue_type="stem"
 /dev_stage="Pooled developmental"
 /note="Vector: Lambda Zap; Contains a mixture of
 internodal stem segments"

BASE COUNT 182 a 124 c 150 g 172 t 2 others
 ORIGIN

Query Match 38.2%; Score 87.4; DB 23; Length 630;
 Best Local Similarity 74.5%; Pred. No. 1e-16;
 Matches 123; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

QY 53 gctcaaccacacagaagtagcttgctgattgctgtaattgagcgaattctggca 112
 Db 92 GCTCAACACACTTACAAGATAAGGCTTGGCGGACTGGCTGTATGGGCCAAATCTCGCA 151

QY 113 ctcaatatgtctgaaaggcttcccaattcccggttaacacaggaaccatttccaaggt 172
 Db 152 CTCATATATGCG-CAACAAGGATCCCAATTCCTGTATATACAGACACATCAAGGT 210

QY 173 tattgggccaataagcagcaacacaggaaggaaaccttcaatt 217
 Db 211 TGACGAAACTGTGGAACGACGACAAACAGGAGGAAATCTTCCACT 255

RESULT 13
 AV410206
 LOCUS
 DEFINITION AV410206 Lotus japonicus young plants (two-week old) Lotus
 japonicus cDNA clone MML068h09_r 5', mRNA sequence.
 ACCESSION AV410206
 VERSION AV410206.1 GI:7723060
 KEYWORDS
 SOURCE EST.
 ORGANISM
 Lotus japonicus
 Lotus japonicus
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Lotus.

REFERENCE
 AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 TITLE Generation of 7137 non-redundant expressed sequence tags from a
 legume, Lotus japonicus
 JOURNAL DNA Res. 7 (2), 127-130 (2000)
 MEDLINE 20277479
 COMMENT Contact: Yasukazu Nakamura
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1..321
 /organism="Lotus japonicus"
 /db_xref="taxon:34305"
 /clone="MML068h09_r"
 /clone_lib="Lotus japonicus young plants (two-week old)"
 /dev_stage="young plants (two-week old)"
 /note="Vector: pBluescriptII SK-; Site_1: ECORI; Site_2:
 XhoI; isolate=Miyakojima MG-20"

BASE COUNT 97 a 79 c 63 g 82 t
 ORIGIN

Query Match 37.4%; Score 85.6; DB 18; Length 321;
 Best Local Similarity 71.6%; Pred. No. 3.1e-16;
 Matches 126; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY 54 ctcaaccacacagaagtagcttgctgattgctgtaattgagcgaattctggca 113
 Db 35 CTCCAACCCCTACAGAATTGGCTTGGCTGTATGGTGCCCAAAACCTTGCTC 94

QY 114 tcaatatgtctgaaaggcttcccaattcccggttaacacaggaaccatttccaaggt 173
 Db 95 TCAACATTGC-TGATAAGGCTTTCGCATTCTGTATATACCAACCAACATCCCAAGGT 153

QY 174 attggccataagcagcaacacaggaaggaaaccttcaatttggggaacaa 229
 Db 154 GATGAGACTCGACGAGCAAAAAGAGAAATCTTCTGTGTATGGCTACCA 209

RESULT 14
 AV410953
 LOCUS
 DEFINITION AV410953 Lotus japonicus young plants (two-week old) Lotus
 japonicus cDNA clone MML200c07_r 5', mRNA sequence.
 ACCESSION AV410953
 VERSION AV410953.1 GI:7740114
 KEYWORDS
 SOURCE EST.
 ORGANISM
 Lotus japonicus
 Lotus japonicus
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 Fabales; Fabaceae; Papilionoideae; Lotus.

REFERENCE
 AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
 TITLE Generation of 7137 non-redundant expressed sequence tags from a
 legume, Lotus japonicus

```

JOURNAL      DNA Res. 7 (2), 127-130 (2000)
MEDLINE      20277479
COMMENT      Contact: Yasukazu Nakamura
              The First Laboratory for Plant Gene Research
              Kazusa DNA Research Institute
              Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
              Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES     source
            1. .379
              /organism="Lotus japonicus"
              /db_xref="taxon:34305"
              /clone="MMW200c07_r"
              /clone_lib="Lotus japonicus young plants (two-week old)"
              /dev_stage="young plants (two-week old)"
              /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
              XhoI; isolate=Miyakojima MG-20"

BASE COUNT   114 a  80 c  89 g  96 t
ORIGIN
Query Match      37.4%; Score 85.6; DB 18; Length 379;
Best Local Similarity 71.6%; Pred. No. 3.2e-16;
Matches 126; Conservative 0; Mismatches 49; Indels 1; Gaps 1;

QY  54 ctcaaccacacagaataggccttggtggtggttaattggcgcaaaatctgcac 113
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   14 CTCAAACCCCTACAGAATTTGGCTTCTGCTGATTTGGCTGTGATGGCCAAACCTTGCTC 73

QY  114 tcaatattgctgaaaggcttcccaattccggttaacaacggaaccatttccaaagtt 173
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   74 TCACATTGC-TGATAAGGCTTTCCCATTTCTGTTATTAAACCGAACCACATCCAAGTT 132

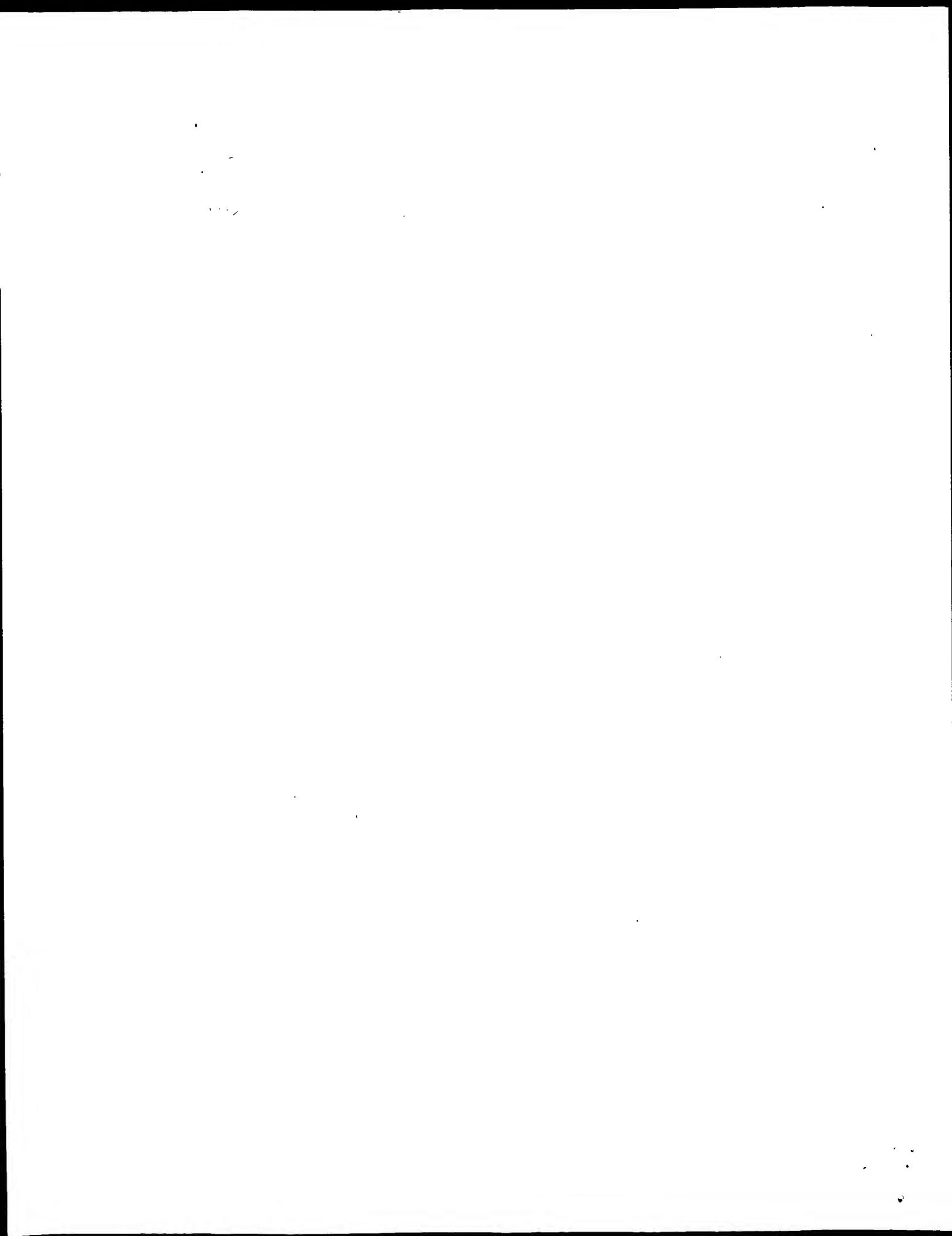
QY  174 attgggcataaagcgaagcaacacggaaggaaccttcaattattggggaacaa 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  133 GATGAGACTGCGAACGACGCAAAAAAAGAGGAATCTTCTGTGTATGGCTACCA 188

RESULT 15
AV425788
LOCUS      AV425788 394 bp mRNA EST 23-MAY-2000
DEFINITION AV425788 Lotus japonicus young plants (two-week old) Lotus
            japonicus cDNA clone MMW059a03_r 5', mRNA sequence.
ACCESSION  AV425788
VERSION     AV425788.1 GI:7784076
KEYWORDS   EST.
SOURCE      Lotus japonicus.
ORGANISM   Lotus japonicus
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1 (bases 1 to 394)
AUTHORS     Asanizu.E., Nakamura.Y., Sato.S. and Tabata.S.
TITLES      Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
JOURNAL     DNA Res. 7 (2), 127-130 (2000)
MEDLINE     20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            -Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.

FEATURES     source
            1. .394
              /organism="Lotus japonicus"
              /db_xref="taxon:34305"
              /clone="MMW059a03_r"
              /clone_lib="Lotus japonicus young plants (two-week old)"
              /dev_stage="young plants (two-week old)"
              /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
              XhoI; isolate=Miyakojima MG-20"

BASE COUNT   115 a  85 c  93 g  101 t
ORIGIN

```



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:30:03 ; Search time 4075.18 Seconds
(without alignments)
281.871 Million cell updates/sec

Title: US-09-300-482-4
Perfect score: 263
Sequence: 1 gaagcacttttgatgtgc.....agnngnacntnnnnnganna 263

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: em_fun:*
- 13: em_hum1:*
- 14: em_hum2:*
- 15: em_in:*
- 16: em_om:*
- 17: em_or:*
- 18: em_ov:*
- 19: em_pat:*
- 20: em_ph:*
- 21: em_pl1:*
- 22: em_ro:*
- 23: em_sts:*
- 24: em_sy:*
- 25: em_un:*
- 26: em_vi:*
- 27: gb_htg1:*
- 28: gb_htg2:*
- 29: gb_in1:*
- 30: gb_in2:*
- 31: em_bal:*
- 32: em_ba2:*
- 33: em_hum3:*
- 34: em_hum4:*
- 35: gb_pr4:*
- 36: gb_htg3:*
- 37: gb_htg4:*
- 38: gb_htg5:*
- 39: gb_htg6:*
- 40: gb_htg7:*
- 41: em_htg1:*
- 42: em_htg2:*
- 43: em_htg3:*

- 44: em_hum5:*
- 45: gb_pl3:*
- 46: gb_pr5:*
- 47: gb_htg8:*
- 48: gb_htg9:*
- 49: gb_htg10:*
- 50: gb_htg11:*
- 51: gb_htg12:*
- 52: gb_htg13:*
- 53: gb_htg14:*
- 54: gb_in3:*
- 55: gb_htg15:*
- 56: gb_htg16:*
- 57: gb_htg17:*
- 58: em_htg4:*
- 59: em_htg5:*
- 60: em_htg6:*
- 61: em_htg7:*
- 62: em_hum6:*
- 63: gb_htg18:*
- 64: gb_htg19:*
- 65: gb_ba3:*
- 66: em_htg8:*
- 67: em_htg9:*
- 68: em_htg10:*
- 69: em_htg11:*
- 70: em_htg12:*
- 71: em_htg13:*
- 72: em_htg14:*
- 73: em_htg15:*
- 74: em_htg16:*
- 75: em_htg17:*
- 76: em_htg18:*
- 77: em_htg19:*
- 78: em_htg20:*
- 79: em_htg21:*
- 80: em_htg22:*
- 81: em_htg23:*
- 82: gb_pr6:*
- 83: gb_pr7:*
- 84: gb_htg20:*
- 85: gb_htg21:*
- 86: gb_htg22:*
- 87: gb_htg23:*
- 88: gb_ro:*
- 89: gb_sts1:*
- 90: gb_sts2:*
- 91: gb_sy:*
- 92: gb_un:*
- 93: gb_vil:*
- 94: gb_vil2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	76.2	29.0	95865	7 AC003970	AC003970 Arabidops
2	68.6	26.1	2058	45 STG6PDHPI	X83923 S.tuberosum
3	66.6	25.3	1396	45 SO000184	AJ000184 Spinacia
4	66.6	25.3	2049	45 SO000182	AJ000182 Spinacia
5	62.2	23.7	1869	45 NTTPG18	AJ001772 Nicotiana
C 6	61.4	23.3	10195	65 U32737	U32737 Haemophilus
7	61	23.2	1630	8 ATG6PDHES	X84229 A.thaliana
8	60.6	23.0	1942	45 NTTPG16	AJ001771 Nicotiana
9	60	22.8	1975	8 ATJ00135	AJ001359 Arabidops
10	59	22.4	1772	8 ATG6PDH15	X84230 A.thaliana
C 11	58.4	22.2	220373	64 ALI58831	ALI58831 Homo sapi
12	56.8	21.6	1976	8 CCA6246	AJ006246 Cyanidium

13	56.6	21.5	463	30	ASU09032	U09032 Anastrepha
14	54	20.5	3015	1	AF269359	AF269359 Staphylococcus aureus
15	54	20.5	3081	1	AF269360	AF269360 Staphylococcus aureus
16	53.6	20.4	1670	29	AB021914	AB021914 Anastrepha
17	52	19.8	35000	45	SPAC9	AL121764 S.pombe chr
18	52	19.8	37013	45	SPAC3A12	Z95395 S.pombe chr
19	51.8	19.7	2049	54	S67872	S67872 Zm-glucose-6-phosphate dehydrogenase
20	51.8	19.7	13551	2	D90906	D90906 Synecocystis sp.
21	51.4	19.5	1808	2	CTU83195	CTU83195 Chlamydia trachomatis
22	51.4	19.5	12333	1	AE001292	AE001292 Chlamydia trachomatis
23	51.4	19.5	16207	1	AE002314	AE002314 Chlamydia trachomatis
24	51	19.4	2750	54	PFAG6PD	M80635 Plasmodium falciparum
25	51	19.4	3244	54	PFLGPH	X74988 P.falciaparum
26	50.6	19.2	2055	3	MRU13899	U13899 Macropus roratus
27	50.6	19.2	2359	7	AF012861	AF012861 Petroselinum
28	50.2	19.1	1862	8	ATHO10970	AJ010970 Arabidopsis thaliana
29	50.2	19.1	10548	1	AE004417	AE004417 Vibrio cholerae
30	49.8	18.9	1690	8	ATHO10971	AJ010971 Arabidopsis thaliana
31	49.8	18.9	2053	45	NTG6PD	X99405 N.tabacum m
32	49.8	18.9	2098	45	STU010712	XJ010712 Solanum tuberosum
33	49.8	18.9	10903	2	D88189	D88189 Actinobacillus
34	49.2	18.7	460	54	CL009033	U09033 Ctenolepis
35	48.8	18.6	1249	1	AF246307	AF246307 Acidithiobacillus
36	48.8	18.6	11895	1	AE001772	AE001772 Thermotoga
37	48.2	18.3	1852	45	NPTCG6	AJ001769 Nicotiana glauca
38	47.6	18.1	2399	45	SCYNL241C	Z71517 S.cerevisiae
39	47.6	18.1	3033	45	SCMET19	X57336 S.cerevisiae
40	47.6	18.1	3782	45	YSG66PD	M34709 S.cerevisiae
41	47.6	18.1	6672	2	NOSFTZOO	L32796 Nostoc sp.
42	47.6	18.1	38855	45	SCCXIV39K	Z69381 S.cerevisiae
43	47.4	18.0	2322	88	MMGPDDH	Z11911 M.musculus
44	47.2	17.9	460	54	RPU09040	U09040 Rhagoletis pomonella
45	46.6	17.7	1821	45	NPTCG9	AJ001770 Nicotiana glauca

Stanford University, 855 California Avenue, Palo Alto, CA 94304,

USA

4 (bases 1 to 95865)

Federspiel,N.A., Palm,C.J., Conway,A.B., Kurtz,D.B., Conway,A.R., Au,M., Araujo,R., Buehler,E., Dewar,K., Feng,J., Kim,C., Li,Y., Oji.O., Osborne,B.I., Shinn,P., Sun,H., Toriumi,M., Vyotskaia,V., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.

Direct Submission

Submitted (28-AUG-1998) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

5 (bases 1 to 95865)

Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A. and Davis,R.W.

Direct Submission

Submitted (26-JUL-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA-94304, USA

On Jun 3, 1998 this sequence version replaced gi:2829860.

Bases 94718-95865 of clone F14J9 overlap with bases 1-1148 of 'IGF' clone F2LM12, gb|AC000132

e-mail for correspondence: arabes@sequence.stanford.edu

Genes with similarity to proteins in the databases are described as 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Grail (informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://compbio.stanford.edu/~chris/GENSCANW.html), Fexa (V.Solovyev & A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark,http://www.cbs.dtu.dk/NetPlantGene.html).

Location/Qualifiers

1..95865

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="I"

/clone="F14J9"

/note="contains phyA molecular marker"

/gene="F14J9.1"

/join(1491..1796,1887..2210,2364..2515,2595..2817)

/note="Similar to rice water stress induced protein gll1537404"

/codon_start=1

/protein_id="AAC33195.1"

/db_xref="GI:3482910"

/translation="MAPEMNKLISYGEKKRVTFTAGTGDYVRGVGLAKRLKRTKS KPLVAVLPDPADHRQLLDQGCVIKEIOPVPDNOFOAMAYVLNVSKUIRW FYEYSKLILYDGDIQVFENIDHLDFDPGNFYAKDCCEKTWSHTQYKLGVCQQCP DYTWPESLGGPKPLPYFNAGVFYPSLYNLLLETLVKVPPTPFAEQDLNNYFK DIKYPPIPVNLVLAMLWRHPENIELNEAKVHYCAAGAKPWRTFGEGNEREDIRK LVKEKWNDIYNDESIDYKNPNVHCQGKEDVHRKPRTLPOFFTDLSEADVLOCAKAPSA "

3405..3971

/gene="F14J9.2"

3405..3971

/gene="F14J9.2"

/note="Hypothetical protein"

/codon_start=1

/protein_id="AAC33196.1"

/db_xref="GI:3482911"

/translation="MIVLVLIYTFTHPIEGVEADGTEALINSICIENEDYGFCNKIIH EKLTPTATIKELTSIIYTTWAHSDTYIFIDNLRKWPGETSLGTCHAVYKRE TNSFLTRFLFSKGEVELMDIEALTAKILECRGDFLIPPYKEPILLEKKRMRLIT MSAYSGHMVKNKASLISVVTAQFNI"

6504..7028

/gene="F14J9.3"

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

6504..7028
 /gene="F14J9.3"
 /note="Hypothetical protein"
 /codon_start=1
 /protein_id="AAC33197.1"
 /db_xref="GI:3482912"
 /translation="MHILSLVLAIIILASVAVAFLEKTDQORVNGICKQTMDFRCS
 SLLINKNTFPASNKEIMNVTVSFAERFAANTFFSTILRNAGDERPDLQACAEAYA
 IVNSAFTKAVTFKQAYSKIVNEKKVSMVADICTDFENVLQINPLTERNQTKI
 LLSMEQIVSHVSS"
 complement(7954..10925)
 /gene="F14J9.4"
 complement(join(7954..8155,8251..8402,8490..8648,
 8762..9014,9101..9217,9294..9356,9577..9606,9895..9938,
 10404..10444,10829..10925))
 /gene="F14J9.4"
 /note="Similar to Mcn21, gi|2598575, Megicaco truncatula
 nodulation induced gene"
 /codon_start=1
 /protein_id="AAC33198.1"
 /db_xref="GI:3482913"
 /translation="MAKSDMLPLAVLVQIGYAGNITSKMAEENDEIIFINIV
 KMVKKNVVKKTKKWCRAIFCQDKTRPKITLRLVQVFCSTIGATGNQVLYFV
 GLQNSPTICATNLPLPAVFLAAIFRQETVGKQSKQAKVIGTIVCVIGAMVLS
 FYHGTIGIGESKHWAYAEINITHGSSGHSNFFLGPPLIMAAVVAWAFIQTGM
 SETFAAPYTTLLMLGSIQCGAIALISDHTISDMSLSPRLFTSALYAGVASALA
 FCLMSAMQKPLYYVSFPLLLVVALFSLWALLEKLYTGTFGMSALYVILGYVL
 WGDREVSKEEREKVKQONHKVSESNEIESRLPVASSNGSTRSTSP"
 12867..15018
 /gene="F14J9.5"
 join(12867..13122,13720..13911,13990..14156,14269..14533,
 14786..15018)
 /gene="F14J9.5"
 /note="Similar to nodulins and lipase; location of EST
 E6C2T7, gi|AA042309, similar to nodulins gi|328240,
 gi|2129854 and others and lipase, gi|2129636"
 /codon_start=1
 /protein_id="AAC33199.1"
 /db_xref="GI:3482914"
 /translation="MATLSHSHSFLVLLPFLILRONLAVAGCCQVPPVFNFGDS
 NSDTGLVAGLVYISGLPNGRSFFORSRLSDGLVDFLCQSLNTSLNPLVLSLV
 GSKFONGFAIVGSGSTLPYVPFALNQLMPLFLFKSALELASIDPLKEMMIGES
 GFNALYMDIQNDIADSFSGKLSRVKLIPNVISEIKSAILYDEGRFVWH
 NNGPLCLQKLSMDYDFSGKSLATYNAAKLFNEGLDMCRDLRLTEKLANIV
 VDIYALKYDLIANSNNGYFEPKPLMACCGYGGPPYNNVNNITCGNGSKSCDEGSRFIS
 WGIHYTETANAIVAMKVLMSQHSPTPTPFHFCGG"
 complement(15279..16596)
 /gene="F14J9.6"
 complement(join(15279..15796,15879..16037,16122..16252,
 16328..16488,16571..16596))
 /gene="F14J9.6"
 /note="Similar to 12-oxophytodienoate reductase,
 gi|2765083 and old-yellow-enzyme homolog, gi|2232254"
 /codon_start=1
 /protein_id="AAC33200.1"
 /db_xref="GI:3482915"
 /translation="MKNEHLTRIVMAPMRYSYGNIPQHVLYYQRTTPGGLLI
 SATGVSETAMAYQNMGIWKEQIEAWKPIVDVAVSHGGIFFQLWHAGRVSHQCO
 PNGESPSTDKPDEADDEFTPPRLRDEIPTIINDFLAARNATEAGFDGVEIH
 GAHGLYLDQFMKDSVNDRTDSYGSLENRCRFAQVTEAVSKETGPRVGRILSPAD
 YMESGDTPKRLGLYMAKSLNRFELVCHMTEPRMKTVSEIFECRESLTPMRAFNGT
 FTVAGSYREDGKAVAGERTDLVAYGRFLFANPDLPRFELNAPLNK"
 17046..19163
 /gene="F14J9.7"
 17046..19163
 /gene="F14J9.7"
 /note="Hypothetical protein; Similar to Arabidopsis
 selenium-binding protein, gi|2244760 and Arabidopsis
 hypothetical proteins gi|3033399, gi|2464864, gi|2244839
 and several others"
 /codon_start=1
 /protein_id="AAC33201.1"
 /db_xref="GI:3482916"

gene

CDS

gene
 complement(join(2043..20480,20574..20774,20850..21052,
 21153..21310,21508..21635,21716..21921,22053..22167,
 22252..22346,22444..22691,22778..22982,23217..23318))
 /gene="F14J9.8"
 /note="Similar to Glucose-6-phosphate dehydrogenases,
 gi|2276344, gi|2829880, gi|2352919 and others; Location of

Query Match

Best Local Similarity 29.08; Score 76.2; DB 7; Length 95865;
 Matches 108; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 1 gaagcacttttgatgttcgctcatgtcttcgaagcagtcctcagaccagagagatgg 60

Db 21876 GAAGCTCTGGTGATGTGGCATGCACCATCTGGAGATAACCTCAGGCACGAGCTGG 21817

Qy 61 aatgcataatttgagaagcatttgcttgatgcactttctccatagctgacga 120

Db 21816 ACTCGAATAAGTAGTGAAGAACCTTTTGGTTTAACTCACATTCGTCCTCAGTTAACA 21757

Qy 121 caatatcttttcaacttcaggaaagcaaatatag 161

Db 21756 AAGTCATCTTCTCTAAGTTTGAAGAGACAAATCTACAG 21716

RESULT 2

STG6PDHPI

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

Asteridae; Solanales; Solanaceae; Solanaceae; Solanum; Potatoe;

section Petota.

1 (bases 1 to 2058)

von Schaeuwen, A., Langenkamper, G., Graeve, K., Wenderoth, I. and

Scheibe, R.

Molecular characterization of the plastidic glucose-6-phosphate

dehydrogenase from potato in comparison to its cytosolic

counterpart

Plant Physiol. 109 (4), 1327-1335 (1995)

MEDLINE

REFERENCE

2 (bases 1 to 2058)

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

2 (bases 1 to 2058)

AUTHORS

TITLE

JOURNAL

FEATURES

source

JOURNAL FEATURES		Unpublished Location/Qualifiers	
source		1..1396	
		/organism="Spinacia oleracea"	
		/plasmid="pZLI"	
		/cultivar="Matador"	
		/db_xref="taxon:3562"	
		/clone="O28FA38"	
		/dev_stage="vegetative"	
		/tissue_type="leaves"	
gene		1..1396	
		/gene="G6PD"	
CDS		1..>1396	
		/gene="G6PD"	
		/EC_number="1.1.1.49"	
		/codon_start=1	
		/product="Glucose-6-phosphate dehydrogenase"	
		/protein_id="CAA03941.1"	
		/db_xref="GI:2276348"	
		/db_xref="SPTREMBL:O24359"	
		/translation="EFSNGHPLNDVSLQNDVAVNPVIAKSIDPSADLQQLPLELVSKVEPTLSIIYVVGASGLAKKRIEPTLPALFVENCLEPENIVFGFSRTMDEELRTMISKTLTCRIDORENGCKEMDHFQRCFYHSGQYNSDDPFGDLCKLKEKAGRLQRLFYLSIPPNIYFVDRVSVSHRASASGWTIVRPFGRDSSSRELFRSFQYLSLSEQIFRIDHMLQILALFAMETPVSLDTERNEKVKVLRSMKPLKQDVVVGVGKHSKGNKSYGTYDDPTVPNNSTVPTFAAALFIDNARWDGVPFLMKAGRALTKRAEIRVQFRVPGNLYKKTFTGTDLDRATNELVLRVQDEAIYLYLKNKVPGLGMLRDLTDNLICYSIYRGEIPDAYERLL"	
misc_feature		959..979	
		/gene="G6PD"	
		/function="G6PD catalytic site"	
BASE COUNT		418 a 252 c 329 g 397 t	
ORIGIN			
Query Match		25.3%; Score 66.6; DB 45; Length 1396;	
Best Local Similarity		52.4%; Pred. No. 3.1e-07;	
Matches 129; Conservative		0; Mismatches 117; Indels 0; Gaps 0;	
Qy	8	tttttgatgttgctcatgtctgtcgaagcagtgctcagaccagcaggaaggaatcgca	67
Db	500	TTTGGATGTTGGTGGATGTTAGCCATCGAGCTTTCTGCAAGTGGATGCACAGG	559
Qy	68	taataattgagaagccatttggtcttgatgcaactttctcccataggtgacacata	127
Db	560	TTATAGTAGAAAAGCTTTTGGACGTGACTCGGATTCTTCTCGTAGCTAAGAGCT	619
Qy	128	ttcttcaaaccttcaggaaaagcaaatatagaaattgancatctactaggaagaa	187
Db	620	TCAAGCAGTACTTGGAGCGAGGACCATATATTCAGGATGACCATTTACTTGGGAGGAGT	679
Qy	188	ncagtnaaatccctncagggtttaaggtttcaaaannnagnttttgagccaccttngann	247
Db	680	TGGTGGAGAAATTTATCTGTTCTAGCTTTTCAATCTAGTTTGAACCTCTATGGTCAA	739
Qy	248	gnacnt 253	
Db	740	GGAAC 745	
RESULT		4	
SO000182		2049 bp mRNA	
LOCUS		Spinacia oleracea mRNA for glucose-6-phosphate dehydrogenase, clones O28FA14 & O30A4.	
DEFINITION		AJ000182	
ACCESSION		AJ000182.1	
VERSION		GI:2276343	
KEYWORDS		G6PD gene; glucose-6-phosphate dehydrogenase.	
SOURCE		Spinacia oleracea	
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; euphyllophytes; Spermatophyta; Magnoliophyta; Magnoliopsida; eudicotyledons; Spinacia oleracea	
REFERENCE		Fink, A.	
AUTHORS		Submitted (21-JUL-1997) Fink A., Laboratoire de Biochimie et de physiologie Vegetale, University of Geneva, 3, Place de l'Universite, CH-1211 Geneve 4, SWITZERLAND	
TITLE		2 (bases 1 to 1396)	
JOURNAL		Fink, A., Diagon, T., Ferroud, P.F., Crespi, P. and Greppin, H. Nucleotide sequences of glucose-6-phosphate dehydrogenase from Spinacia oleracea cloned by RT-PCR	
23-JUL-1997			
/tissue_type="green leaves"			
/cell_type="mesophyll"			
/clone_lib="lambda ZAPII"			
/clone="pBSK-plafull #3, #4"			
41..1774			
/codon_start=1			
/product="glucose-6-phosphate dehydrogenase"			
/protein_id="CAA58775.1"			
/db_xref="GI:1197385"			
/db_xref="SWISS-PROT:Q43839"			
/translation="MGVQLRLNFCSSSSAATSPSTFNGTTPYFCKRFNLPFRTOPLN WWSGIYRIOPRKHFEVFSNGPFLNARSVDQVQVPLTELGSDDTFTVSLVIGASGDL AKKILPALFALFYEDDFENFVFGYSRTKLKKEGCRVSNRFLYLSIPNIFVDVVRCAK MEHLEFCFYHQYINSEDDFAELDKLKEGCRVSNRFLYLSIPNIFVDVVRCAK LKASFTGWTRIVIVKEPFGDLESSELTRSLKKYLTTEEQIFRIDHMLQILALF LRFSLVPELWSNRYIRNVQVIFSEDFTEGGRGYFDHYGIIIRDIMQHLLQILALF AMETPVSLDAEDIRNEKVKVLRSMRPLQLEDVLGOYKHSKSAKSPAYATDPTVFN GSITPTFAALFIDNARWDGVPFLMKAGRALTKRAEIRVQFRVPGNLYKRNFTCD MDKATNELVLRQDEAIYLYLKNKVPGLGMLRDLRDSNLNLYKAKYRGEIPDAYERLL LDATGEGRRLLFIRSELDAAWALFTPLLKLEBKKAPELIPYGSRGVPVGAHYLAANK NVRWGLSGDD"			
BASE COUNT		606 a 372 c 454 g 626 t	
ORIGIN			
Query Match		26.1%; Score 68.6; DB 45; Length 2058;	
Best Local Similarity		52.8%; Pred. No. 9.5e-08;	
Matches 131; Conservative		0; Mismatches 117; Indels 0; Gaps 0;	
Qy	6	acttttgatgttgctcatgtctgtcgaagcagtgctcagaccagcaggaaggaatcg	65
Db	667	ATTCGTGGATGTTGGTGGATGTTGCAAGCTTTAAAGCTTCTTCAACAAGTGGTGACAA	726
Qy	66	cataattattgagaagccatttggtcttgatgcaactttctcccataggtgacacata	125
Db	727	GGTCATGTTTGAGAAACCATTTGGTCTGATTTAGAAATCATCTAGTAGCTGACCAGATC	786
Qy	126	tcttcttcaaaccttcaggaaaagcaaatatagaaattgancatctactaggaagaa	185
Db	787	CCTAAAGAGTACCTAACAGGAGGACCAATTTCCGAATTGACCAATTTACTTGGGAAAGA	846
Qy	186	tcncagtnaaatccctncagggtttaaggtttcaaaannnagnttttgagccaccttngag	245
Db	847	ACTTGTGAGATCTCTAGTCTTCGGTCTCGAATCTTGCTTTGAGCCCTCTTTGGTC	906
Qy	246	nnnnaent 253	
Db	907	CAGAAACT 914	
RESULT		3	
SO000184		1396 bp mRNA	
LOCUS		Spinacia oleracea mRNA for glucose-6-phosphate dehydrogenase, partial, clone O28FA38.	
DEFINITION		AJ000184	
ACCESSION		AJ000184.1	
VERSION		GI:2276347	
KEYWORDS		G6PD gene; glucose-6-phosphate dehydrogenase.	
SOURCE		Spinacia oleracea	
ORGANISM		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.	
REFERENCE		Fink, A.	
AUTHORS		Submitted (21-JUL-1997) Fink A., Laboratoire de Biochimie et de physiologie Vegetale, University of Geneva, 3, Place de l'Universite, CH-1211 Geneve 4, SWITZERLAND	
TITLE		2 (bases 1 to 1396)	
JOURNAL		Fink, A., Diagon, T., Ferroud, P.F., Crespi, P. and Greppin, H. Nucleotide sequences of glucose-6-phosphate dehydrogenase from Spinacia oleracea cloned by RT-PCR	
23-JUL-1997			

[illegible]

```
/product="H. influenzae predicted coding region HI0554"
/protein_id="AAC22217.1"
/db_xref="GI:1573540"
/translation="MTKYAGIHRRSIRLKHYNRSEGFYFITIOCKNKECLFGHI
INQOLNDLNGYKQWENPMFFPQVRIDFVPMNHLHIIETIEQVKGKCNPL
QLRATQLPKGTSGTIGSIVRRFRAGVTSWAKNSEIFDVQWRNYEHLIRDEKSYLQ
IYEYQNNPILWEOQLYVD"
complement(4243..4482)
/gene="HI0555"
CDS
complement(4243..4482)
/gene="HI0555"
/notes="similar to GB:M87483 SP:Q02009 PID:551879 percent
identity: 40.21; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC22211.1"
/db_xref="GI:1573541"
/translation="MEILAVILTALVTLEHFYILYLEMFTIESKSGVRQSRVQLRFG
SRDVFWLCHOPNFGDLFLRLCDYRRSFWGSYDEE"
complement(4482..4973)
/gene="HI0556"
CDS
complement(4482..4973)
/notes="similar to SP:P46016 PID:556607 percent identity:
35.82; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, putative (devb)"
/protein_id="AAC22212.1"
/db_xref="GI:1573542"
/translation="MVPPSPESNYGEVOKLLFDHILQIPAEHNRHNGENHPELKR
FEELSAVINGVEDMILLGMGDIGHTASLPQHTNFDENLAVIAKHPSGQIRLSK
TAKLIEQAKRITLVGESKADILKIQTPAENLPYPAKIRAKNGVTEWYLDKAAV
RL"
complement(5059..5181)
/gene="HI0557"
CDS
complement(5059..5181)
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="H. influenzae predicted coding region HI0557"
/protein_id="AAC22218.1"
/db_xref="GI:1573547"
/translation="MNYISEPTAQHAVDKIAQEFVIYSQNLNHSRTYFFPWFDA"
complement(5264..6748)
/gene="HI0558"
CDS
complement(5264..6748)
/notes="similar to SP:P77809 PID:1651208 percent identity:
85.22; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase (zwf)"
/protein_id="AAC22213.1"
/db_xref="GI:1573543"
/translation="MOTDNCNIVFCASDLTHRLKLIPLALNLYKIGRLSENFSLVG
ARSDLNDEFREKMRALHNEETIPETLDAFCASHLYOAVNTSDAQYKGLVPRILD
LHDKYQTCNTFTYMSPPSLYGVIPCELAHGLNTEYGNKRIIVKPFYDEKTAQ
TLIDQYHREFFEQHYRIDYLGKTVQNLVLRFSNGWFEPLNRRNFIDVVEITGAE
SIGVERGGYDGSAMRDMFQNHLLQVLAMVEPPAIINANSRMRDEKAVKMSLRP
LTSDEMNNKRLVQYTAETNGKMEKYLEEGVPAANSRTETYIALRCEIENWRAGV
PFYVTRTPARVTEIVTHKTPHPVFSQNAPEKLLIIRIQDEALSMRFGLUKPGV
ACFEAKVSMDFRYADLAGAQLVTAYERLLDLMKMGDTLFRITADVAHAANKFQVPII
DYKANGRIHEAEATGWGPVADKIAKQGVKWRKPSGLMKKV"
complement(6824..7510)
/gene="HI0559"
CDS
complement(6824..7510)
/notes="similar to SP:P22255 GB:M55170 GB:M80795 PID:145275
PID:145690 percent identity: 36.07; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sulfite synthesis pathway protein (cysQ)"
/protein_id="AAC22214.1"
/db_xref="GI:1573544"
/translation="MTEADLFVSQFLTEKLTALFPNVPVLSBENCHISFEERKNWKEY
Query Match 23.3%; Score 61.4; DB 65; Length 10195;
Best Local Similarity 52.1%; Pred. No. 6.7e-06;
Matches 110; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 53 agggatggaatcgcataatatttgagaagccatttgcttgatgcactttcccata 112
DB 6327 ATGCTGGAACGATATTATCGTTGAAAACCTTTTGGTTACGATGAAAAACACGCAAA 6368
QY 113 ggcgcacacatatctcttcacatttcaggaaagcaaatatagataatgancatc 172
DB 6267 CATTAGACGTACAAATCCACCGTTCTTGAAGAACACCAAAATTTATCGTATCGATCAT 6208
QY 173 tactagaaggaaatcncagtnaaatcctcncaggttaagggtttcaaaannagntttg 232
DB 6207 ATTTAGTAAGAAACCGTTCAAACTGCTCGTATACGTTTTTCAAAATGGTTGGTTTG 6148
QY 233 agccacttngagngnactnnnnnganna 263
DB 6147 AACCGCTTTGGAACCGTAATTTTCATTGATTA 6117
RESULT 7
LOCUS ATG6PDHE5 1630 bp mRNA PLN 31-JAN-1996
DEFINITION A.thaliana mRNA for glucose-6-phosphate dehydrogenase (clone E5).
ACCESSION X84229
VERSION X84229.1 GI:1166404
KEYWORDS 96ph gene; glucose-6-phosphate dehydrogenase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1630)
AUTHORS Fink,A., Greppin,H. and Tacchini,P.
TITLE Nucleotide sequence of a cDNA encoding the glucose-6-phosphate
dehydrogenase from Arabidopsis thaliana
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1630)
AUTHORS Fink,A.F.A.
TITLE Direct Submission
JOURNAL
FEATURES
source
1. 1630
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/dev_stage="adult"
/tissue_type="leaves, stems, some flowers and roots"
/clone_lib="lambda ZAP"
/clone="E5"
<1..1630
/gene="G6PDH"
1..1630
/gene="G6PDH"
<1..1480
/gene="G6PDH"
/EC_number="1.1.1.49"
/codon_start=2
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="CAA59011.1"
CDS
/notes="similar to SP:P22255 GB:M55170 GB:M80795 PID:145275
PID:145690 percent identity: 36.07; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sulfite synthesis pathway protein (cysQ)"
/protein_id="AAC22214.1"
/db_xref="GI:1573544"
/translation="MTEADLFVSQFLTEKLTALFPNVPVLSBENCHISFEERKNWKEY
Query Match 23.3%; Score 61.4; DB 65; Length 10195;
Best Local Similarity 52.1%; Pred. No. 6.7e-06;
Matches 110; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 53 agggatggaatcgcataatatttgagaagccatttgcttgatgcactttcccata 112
DB 6327 ATGCTGGAACGATATTATCGTTGAAAACCTTTTGGTTACGATGAAAAACACGCAAA 6368
QY 113 ggcgcacacatatctcttcacatttcaggaaagcaaatatagataatgancatc 172
DB 6267 CATTAGACGTACAAATCCACCGTTCTTGAAGAACACCAAAATTTATCGTATCGATCAT 6208
QY 173 tactagaaggaaatcncagtnaaatcctcncaggttaagggtttcaaaannagntttg 232
DB 6207 ATTTAGTAAGAAACCGTTCAAACTGCTCGTATACGTTTTTCAAAATGGTTGGTTTG 6148
QY 233 agccacttngagngnactnnnnnganna 263
DB 6147 AACCGCTTTGGAACCGTAATTTTCATTGATTA 6117
RESULT 7
LOCUS ATG6PDHE5 1630 bp mRNA PLN 31-JAN-1996
DEFINITION A.thaliana mRNA for glucose-6-phosphate dehydrogenase (clone E5).
ACCESSION X84229
VERSION X84229.1 GI:1166404
KEYWORDS 96ph gene; glucose-6-phosphate dehydrogenase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1630)
AUTHORS Fink,A., Greppin,H. and Tacchini,P.
TITLE Nucleotide sequence of a cDNA encoding the glucose-6-phosphate
dehydrogenase from Arabidopsis thaliana
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1630)
AUTHORS Fink,A.F.A.
TITLE Direct Submission
JOURNAL
FEATURES
source
1. 1630
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/dev_stage="adult"
/tissue_type="leaves, stems, some flowers and roots"
/clone_lib="lambda ZAP"
/clone="E5"
<1..1630
/gene="G6PDH"
1..1630
/gene="G6PDH"
<1..1480
/gene="G6PDH"
/EC_number="1.1.1.49"
/codon_start=2
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="CAA59011.1"
CDS
/notes="similar to SP:P22255 GB:M55170 GB:M80795 PID:145275
PID:145690 percent identity: 36.07; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sulfite synthesis pathway protein (cysQ)"
/protein_id="AAC22214.1"
/db_xref="GI:1573544"
/translation="MTEADLFVSQFLTEKLTALFPNVPVLSBENCHISFEERKNWKEY
```


/db_xref="GI:1166405"
 /translat="SQSTVSTIVVVGASGLDRAKKIPFALFYEGCLPEHFTVGY
 SRSMTEVLRNMTLTCRDKRAMCKEFLKRCFYISQYDSQEHFELDCK
 LKEHAGRIENRFLYILIPNIVDAVKCASTASSVNGWTRVIVEKPEGROSETSA
 LTKSLQYLEDQIFRHDYVLGKELVENLVLRFSLIFEPLWSROYIRNVOIFSED
 FTGREGYFDNTGIIINDIMONHLLQILALFAMETPVSLDAEDIRNEKVKVLRSMPI
 RVEDVIGQYKSTGKGLTPAYTDKTVPGSLTPFAAALFIDNARWDGVPFLMK
 AGKALHTRSAEIRQFRHVPNGNLYNRNTGSDLDQATNELVIRVOPDEAFYKLNKVP
 GLGRLNLLISARYSKEIPDAYERLLDAIEGERLFIIRSDLEDAWLSLFTPL
 LKTEKKRIPEYYPYGRGPGVGAHYLAARHKVQWGDVSDQ"
 /gene="G6PDH"
 1600..1630
 /db_xref="G6PDH"
 1630

polyA_signal
 polyA_site
 BASE COUNT 489 a 292 c 366 g 483 t
 ORIGIN

Query Match 23.28; Score 61; DB 8; Length 1630;
 Best Local Similarity 52.1%; Pred. No. 8.4e-06;
 Matches 124; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
 QY 6 acttttgatgtgcgtcattcttgcagcagtgctcagaccagaggaaggaatcg 65
 Db 373 ATTGTAGATGCTGTAATGTCAGACATCTCTCTCATCTGTCATGATGGACTAG 432
 QY 66 cataatatttgagaagccatttgggttgatgacattcttcccatagctgacacaata 125
 Db 433 GGTATGTCGAGAAACCTTTTGGTCGAGATTCTGAAACCTCGGCTGCTTAAACGAAATC 492
 QY 126 tcttcttcaacttccagaaaagcaaatatagaaattgancatctactaggaagaa 185
 Db 493 CCTCAAGCAGTATTGGAGAGAGATCAAAATTTTAGTAGTAGACCATACCTAGGAAGGA 552
 QY 186 tncagtnaaatcctcaggtttaaggggtttcaannnagnttttgagccacttng 243
 Db 553 GCTAGTCGAGACTATCTGTTCTTCGATCTCAACCTTATATTGAGCGCTATGG 610

RESULT 8
 LOCUS NTTPG16 1942 bp mRNA PLN 17-AUG-1999
 DEFINITION Nicotiana tabacum mRNA for plastidic glucose-6-phosphate dehydrogenase TPGL6.
 ACCESSION AJ001771
 VERSION AJ001771.1 GI:3650343
 KEYWORDS glucose-6-phosphate dehydrogenase.
 SOURCE common tobacco.
 ORGANISM Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 (bases 1 to 1920)
 AUTHORS Wendt,U.K., Hauschild,R., Lange,C., Pietersma,M., Wenderoth,I. and von Schaeuwen,A.
 TITLE Evidence for functional convergence of redox regulation in G6PDH isoforms of cyanobacteria and higher plants
 JOURNAL Plant Mol. Biol. 40 (3), 487-494 (1999)
 MEDLINE 99364543
 REFERENCE 2 (bases 1 to 1942)
 AUTHORS von Schaeuwen,A.
 TITLE Direct Submission
 JOURNAL Submitted (18-SEP-1997) von Schaeuwen A., Plant Physiology, University Osnabrueck, Barbarastr. 11, D-49069 Osnabrueck, GERMANY revised by [3]
 REFERENCE 3 (bases 1 to 1942)
 AUTHORS von Schaeuwen,A.
 TITLE Direct Submission
 JOURNAL Submitted (23-SEP-1998) vonSchaeuwen A., Plant Physiology, University Osnabrueck, Barbarastr. 11, D-49069 Osnabrueck, GERMANY On Sep 25, 1998 this sequence version replaced gi:3021530.

FEATURES

Source

Location/Qualifiers

1..1942
 /organism="Nicotiana tabacum"
 /cultivar="Samsun NN"
 /db_xref="taxon:4097"
 /cell_type="mesophyll"
 /tissue_type="young sink leaves"
 sig_peptide 41..232
 mRNA 41..1801
 /note="TPGL6"
 /product="glucose-6-phosphate dehydrogenase"
 233..1798
 /product="glucose-6-phosphate dehydrogenase"
 mat_peptide 584 a 353 c 442 g 563 t
 BASE COUNT
 ORIGIN

Query Match 23.0%; Score 60.6; DB 45; Length 1942;
 Best Local Similarity 50.8%; Pred. No. 1.1e-05;
 Matches 126; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
 QY 6 acttttgatgtgcgtcattcttgcagcagtgctcagaccagaggaaggaatcg 65
 Db 694 ATTCTGTGATGCTGGTGAAGTGTGCAAGTCTTAAAGCTTCTTCAACAGTGGGTGGACAAG 753
 QY 66 cataatatttgagaagccatttgggttgatgacattcttcccatagctgacacaata 125
 Db 754 AGTCATGTTGAGAAACCGTTTGGTCGTGACTTGAATCATCTAGCGAGCTGACCAGATG 813
 QY 126 tcttcttcaacttccagaaaagcaaatatagaaattgancatctactaggaagaa 185
 Db 814 CCTAAGAAGTACTACTGAGGAGCAAAATATTCGCAATTTGACCATTTACTTTGGGAAGGA 873
 QY 186 tncagtnaaatcctcaggtttaaggggtttcaannnagnttttgagccacttng 245
 Db 874 ACTGTGTTGAGAACTCTCCGTTCTCTGTTCTCGAATCTTGTCTTTCGAGCCTCTGTGTC 933
 QY 246 mngnact 253
 Db 934 CAGAAACT 941

RESULT 9

ATJ00135

LOCUS

DEFINITION

ATJ00135 1975 bp mRNA PLN 17-AUG-1999
 Arabidopsis thaliana cDNA encoding plastidic glucose-6-phosphate dehydrogenase.

ACCESSION

AJ001359.1 GI:3021304

VERSION

glucose-6-phosphate dehydrogenase.

KEYWORDS

thale cress.

SOURCE

Arabidopsis thaliana

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 1975)
 AUTHORS Wendt,U.K., Hauschild,R., Lange,C., Pietersma,M., Wenderoth,I. and von Schaeuwen,A.
 TITLE Evidence for functional convergence of redox regulation in G6PDH isoforms of cyanobacteria and higher plants
 JOURNAL Plant Mol. Biol. 40 (3), 487-494 (1999)
 MEDLINE 99364543
 REFERENCE 2 (bases 1 to 1975)
 AUTHORS von Schaeuwen,A.
 TITLE Direct Submission
 JOURNAL Submitted (18-SEP-1997) von Schaeuwen A., Plant Physiology, University Osnabrueck, Barbarastr. 11, D-49069 Osnabrueck, GERMANY

JOURNAL

Plant Mol. Biol. 40 (3), 487-494 (1999)

MEDLINE

99364543

AUTHORS

von Schaeuwen,A.

TITLE

Direct Submission

JOURNAL

Submitted (18-SEP-1997) von Schaeuwen A., Plant Physiology, University Osnabrueck, Barbarastr. 11, D-49069 Osnabrueck, GERMANY

FEATURES

source

1..1975
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /tissue_type="Above ground parts"

CDS

```

71. 1801
/codon_start=1
/product="plastidic glucose-6-phosphate dehydrogenase"
/protein_id="CAA04696.1"
/db_xref="GI:3021305"
/db_xref="SPTREMBL:O65577"
/translation="MATHSMIIPSPSSSSSLATAAPPKETLPLFSRLTPFRKSLF
SOVLRPFPAEKHSOLDTSNCAINFAQLSDGDLTEHVTKGESTLSITVVGASGDL
AKKIFLPALEFTEGCLPQDFSVFYARTKLTHEELRDMISSTLCRIDQREKCGDK
MEQKLCFYHSGYNSSEDFALNKKLKEKAGKISNRLYILSPNLFVDVVRCAASLRAS
LRASNGWTRVIVKEPFGDSESSGELTRCLQYLTEQIFRIDHYLKGELVENLSV
LRFNLVPELWSNRYIRNVQLIFSEDFTEGRGVDYGIIRDIMQHLLQILALF
AMETPSVLDADIIRSEKVKVLRSMKPLRLDEDVVGQYKGNHKGKTYPTDDPTVFN
HSLPTFAAAMEINNAWDGVPFLMKAGKALHTRGAEIRVQFRHVPGLNLYKSFATN
LDNATNELVIRVQDEGIYLRINNKVPGLMRLDRDLNLLYRSYRPREIPDAYERLL
LDALAEARRLFIKSELDAAWDLFTALKELEKTIPELYPYGSRGPVGAHYLASKY
NVRWGDLGEA"
BASE COUNT 583 a 382 c 447 g 563 t
ORIGIN

```

```

Query Match 22.8%; Score 60; DB 8; Length 1975;
Best Local Similarity 52.5%; Pred. No. 1.5e-05;
Matches 114; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 37 agtgcagaccagaggggaatcgcaataatttgcagagccatttgcttgat 96
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 AGAGCCTCCTCAGAGATGGCTGACACAGGCTATTCTGGAAGCCATTGGCCGTGAC 787

QY 97 gcaatttctccatagctgacacaatatcttcttcaaaccttcaggaaagcaata 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 TCTGAATCACTCGGAGAGTAACTAGATGCTGAAACAGTATCTACAGAGGAGCAATC 847

QY 157 tatgaattgancatctactaggaagaatcncagtnaaatcctcaggtttaagggtt 216
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 848 TTCAGATTGATCATCTATTTCGGAAAGGAGCTGTGTGAGAACCTTTCAGTCTCCGATTC 907

QY 217 tcaaanngnntttgagccacttngagngnnaact 253
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 908 TCAAACTCTGTTTCGAGCCTCTGTGTCAGAAAT 944

```

```

RESULT 10
LOCUS ATG6PDH15 1772 bp mRNA PLN 15-AUG-1997
DEFINITION A.thaliana mRNA for glucose-6-phosphate dehydrogenase (Clone E15).
ACCESSION X84230
VERSION 84230.1 GI:1174335
KEYWORDS 6pdh gene; glucose-6-phosphate dehydrogenase.
SOURCE thale cress, thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Capparales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1772)
Fink, A.F.A.
Direct Submission
Submitted (23-JAN-1995) A.F.A. Fink, Lab de Biochimie et de
Physiologie, Vitales, Universiti de Geneve, 3 Place de
1/Universiti, 1211 Geneva, SWITZERLAND
2 (bases 1 to 1772)
Fink, A., Greppin, H. and Tacchini, P.
Nucleotide sequence of a cDNA encoding the glucose-6-phosphate
dehydrogenase (GenBank X84230) from Arabidopsis thaliana
Plant Physiol. 108, 1343-1343 (1995)
Ref [3]: Plant Gene Register PGR95-021 (1995).
Location/Qualifiers
1. 1..1772
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/haplotype="diploid"
/dev_stage="adult"

```

```

REFERENCE
AUTHORS Fink, A.F.A.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-1995) A.F.A. Fink, Lab de Biochimie et de
Physiologie, Vitales, Universiti de Geneve, 3 Place de
1/Universiti, 1211 Geneva, SWITZERLAND
2 (bases 1 to 1772)
Fink, A., Greppin, H. and Tacchini, P.
Nucleotide sequence of a cDNA encoding the glucose-6-phosphate
dehydrogenase (GenBank X84230) from Arabidopsis thaliana
Plant Physiol. 108, 1343-1343 (1995)
COMMENT Ref [3]: Plant Gene Register PGR95-021 (1995).
FEATURES
source Location/Qualifiers
1. 1..1772
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/haplotype="diploid"
/dev_stage="adult"

```

```

/tissue_type="leaves, stems, flowers and roots"
/cell_line="A.thaliana Co"
/clone_lib="lambda ZAP"
/clone="E15"
141..1685
/gene="G6PDH"
141..1685
/gene="G6PDH"
141..1685
/gene="G6PDH"
/EC_number="1.1.1.49"
/codon_start=1
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="CAA59012.1"
/db_xref="GI:1174336"
/db_xref="SWISS-PROT:Q43727"
/translation="MGVATNFASLDQSGDLTEHVTKGESTLSITVVGASGLAKKK
IFPDLFALFYEGCLPQDFSVFYARTKLTHEELRDMISSTLCRIDQREKCGDKMEQF
LKRCFYHSGYNSSEDFALNKKLKEKAGKISNRLYILSPNLFVDVVRCAASLRAS
SENGWTRVIVKEPFGDSESSGELTRCLQYLTEQIFRIDHYLKGELVENLSVLRFS
NLVPELWSNRYIRNVQLIFSEDFTEGRGVDYGIIRDIMQHLLQILALFAMET
PSVLDADIIRSEKVKVLRSMKPLRLDEDVVGQYKGNHKGKTYPTDDPTVFNHSLT
PTFAAAMEINNAWDGVPFLMKAGKALHTRGAEIRVQFRHVPGLNLYKSFATNLDNA
TNELVIRVQDEGIYLRINNKVPGLMRLDRDLNLLYRSYRPREIPDAYERLLDAI
EGERLFIKSELDAAWDLFTALKELEKTIPELYPYGSRGPVGAHYLASKYNVRW
GDLGEA"
polyA_signal 1756..1761
polyA_signal 1772
BASE COUNT 526 a 331 c 419 g 496 t
ORIGIN

```

```

Query Match 22.4%; Score 59; DB 8; Length 1772;
Best Local Similarity 50.4%; Pred. No. 2.8e-05;
Matches 125; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 6 acttttgatgtgcgtcattgttcagagcagtcgtcagaccagagggatggaatcg 65
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 ATTCGTTGATGTGGTTAGGTGTGCAAGTCTTAGAGCCTCTCAGAGAAATGGTGCAAG 640

QY 66 cataatatattgagaagccattgttgctgacacttcttcccataggtgacacaata 125
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641 GGTCATTGTGGAAAGCCATTGGTCTGCTGACTGTAATCCGGAGAGTTAACTAGATG 700

QY 126 tcttcttcaaaccttcaggaaagcaataatatagaaattgancatctactaggaagaa 185
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 TCTGAACACGATCTTACAGAGGAGCAAACTCTTCAGGATTGATCACTATTTCGGAAAGGA 760

QY 186 tencagtnaaatcctcaggtttaagggtttcaaanngnntttgagccacttngag 245
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 761 GCTTGTGAGACCTTTCAGTCTCGGATTCCTCAAACTCTTGTTCGAGCCTCTGTGTC 820

QY 246 nngnaact 253
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 821 AAGAAAT 828

```

```

RESULT 11
LOCUS AL158831/C
DEFINITION Homo sapiens chromosome 9 clone RP11-564A4, *** SEQUENCING IN
PROGRESS ***, 33 unordered pieces.
ACCESSION AL158831
VERSION AL158831.5 GI:9212620
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 220373)
Plumb, B.
Direct Submission
Submitted (21-MAY-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

```

Sat Nov 4 18:11:37 2000

COMMENT

On Jul 15, 2000 this sequence version replaced gi:7939701.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: hunquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA564A4
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 200906 bases at least Q40
 Consensus quality: 208868 bases at least Q30
 Consensus quality: 213166 bases at least Q20
 Insert size: 217173; sum-of-contigs
 Insert size: 157787; agarose-fp
 Quality coverage: 3.42x in Q20 bases; sum-of-contigs Quality
 coverage: 4.71x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 33 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 19404: contig of 19404 bp in length
 * 19405 19504: gap of 100 bp
 * 19505 20573: contig of 1069 bp in length
 * 20574 20673: gap of 100 bp
 * 20674 22382: contig of 1709 bp in length
 * 22383 22482: gap of 100 bp
 * 22483 30989: contig of 8507 bp in length
 * 30990 31089: gap of 100 bp
 * 31090 34726: contig of 3637 bp in length
 * 34727 34826: gap of 100 bp
 * 34827 36278: contig of 1452 bp in length
 * 36279 36378: gap of 100 bp
 * 36379 57138: contig of 20760 bp in length
 * 57139 57238: gap of 100 bp
 * 57239 61395: contig of 4157 bp in length
 * 61396 61495: gap of 100 bp
 * 61496 63183: contig of 1888 bp in length
 * 63184 63283: gap of 100 bp
 * 63284 64346: contig of 1063 bp in length
 * 64347 64446: gap of 100 bp
 * 64447 66218: contig of 1772 bp in length
 * 66219 66318: gap of 100 bp
 * 66319 67584: contig of 1266 bp in length
 * 67585 67684: gap of 100 bp
 * 67685 68875: contig of 1191 bp in length
 * 68876 68975: gap of 100 bp
 * 68976 71548: contig of 2573 bp in length
 * 71549 71648: gap of 100 bp
 * 71649 73089: contig of 1441 bp in length
 * 73090 73189: gap of 100 bp
 * 73190 81026: contig of 7837 bp in length
 * 81027 81126: gap of 100 bp
 * 81127 82976: contig of 1850 bp in length
 * 82977 83076: gap of 100 bp
 * 83077 88550: contig of 5474 bp in length
 * 88551 88650: gap of 100 bp
 * 88651 89972: contig of 1322 bp in length
 * 89973 90072: gap of 100 bp
 * 90073 94252: contig of 4180 bp in length
 * 94253 94352: gap of 100 bp
 * 94353 97066: contig of 2714 bp in length
 * 97067 97166: gap of 100 bp
 * 97167 100212: contig of 3046 bp in length
 * 100213 100312: gap of 100 bp
 * 100313 103147: contig of 2835 bp in length

* 103148 103247: gap of 100 bp
 * 103248 104757: contig of 1510 bp in length
 * 104758 104857: gap of 100 bp
 * 104858 106680: contig of 1823 bp in length
 * 106681 106780: gap of 100 bp
 * 106781 108222: contig of 1442 bp in length
 * 108223 108322: gap of 100 bp
 * 108323 109526: contig of 1204 bp in length
 * 109527 109626: gap of 100 bp
 * 109627 156013: contig of 46387 bp in length
 * 156014 156113: gap of 100 bp
 * 156114 162589: contig of 6476 bp in length
 * 162590 162689: gap of 100 bp
 * 162690 189785: contig of 27096 bp in length
 * 189786 189885: gap of 100 bp
 * 189886 198960: contig of 9075 bp in length
 * 198961 199060: gap of 100 bp
 * 199061 211582: contig of 12522 bp in length
 * 211583 211682: gap of 100 bp
 * 211683 220373: contig of 8691 bp in length.
 * * * * *
 FEATURES
 Location/Qualifiers
 1..220373
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-564A4"
 /clone_lib="RPC1-11.2"
 1..19404
 /note="assembly_fragment:01468"
 fragment_chain:1
 clone_end:SP6
 vector_side:left
 19505..20573
 /note="assembly_fragment:00210"
 fragment_chain:1
 20674..22382
 /note="assembly_fragment:00702"
 fragment_chain:2
 22483..30989
 /note="assembly_fragment:01191"
 fragment_chain:2
 31090..34726
 /note="assembly_fragment:00723"
 fragment_chain:2
 34827..36278
 /note="assembly_fragment:00747"
 fragment_chain:3
 36379..57138
 /note="assembly_fragment:00108"
 fragment_chain:3
 57239..61395
 /note="assembly_fragment:01455"
 fragment_chain:3
 61496..63183
 /note="assembly_fragment:01321"
 fragment_chain:4
 63284..64346
 /note="assembly_fragment:01175"
 fragment_chain:4
 64447..66218
 /note="assembly_fragment:00029"
 66319..67584
 /note="assembly_fragment:00187"
 /note="assembly_fragment:00187"
 67685..68875
 /note="assembly_fragment:00310"
 68976..71548
 /note="assembly_fragment:00352"
 71649..73089
 /note="assembly_fragment:00419"
 73190..81026
 /note="assembly_fragment:00466"
 81127..82976
 /note="assembly_fragment:00799"

misc_feature 83077..88550
 /note="assembly_fragment:00825"
 misc_feature 88651..89972
 /note="assembly_fragment:00990"
 misc_feature 90073..94252
 /note="assembly_fragment:01020"
 misc_feature 94353..97066
 /note="assembly_fragment:01043"
 misc_feature 97167..100212
 /note="assembly_fragment:01173"
 misc_feature 100313..103147
 /note="assembly_fragment:01212"
 misc_feature 103248..104757
 /note="assembly_fragment:01234"
 misc_feature 104858..106680
 /note="assembly_fragment:01312"
 misc_feature 106781..108222
 /note="assembly_fragment:01421"
 misc_feature 108323..109526
 /note="assembly_fragment:01462"
 misc_feature 109627..156013
 /note="assembly_fragment:00918"
 misc_feature 156114..162589
 /note="assembly_fragment:00811"
 misc_feature 162690..189785
 /note="assembly_fragment:00625"
 misc_feature 189886..198960
 /note="assembly_fragment:00180"
 misc_feature 199061..211582
 /note="assembly_fragment:00696"
 misc_feature 211683..220373
 /note="assembly_fragment:01298"
 fragment_chain:5

Query Match 22.2%; Score 58.4; DB 64; Length 220373;
 Best Local Similarity 52.6%; Pred. No. 4e-05;
 Matches 110; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
 QY 45 gaccacagggatggaatgcacataatcttcttcaaaactttcagaaagccatttggcttgcacattc 104
 DB 91224 GAACGTAAAGGGTTTAAACCAATATATAGAAAACCATTTGGTGATGATTTAAATC 91165
 QY 105 ttccataggtgcacacataatcttcttcaaaactttcagaaagccatttggcttgcacattc 164
 DB 91164 TGCAGAAAATTAATTAATCAATAGAAAGCTCTTTTAAAGAGAGAAATATTTAGAAT 91105
 QY 165 tgancatctactaggaaggaatcncagtnaaatcctcncaggtttaaagggtttcacaann 224
 DB 91104 TGATCATTATTTAGGTAAAGAAATGATTCAAATATATGACGCTTCGATTTGGAAACAC 91045
 QY 225 agntttgagccacttngagngnact 253
 DB 91044 TATATTGAACCACTTTGGAAATAAAGT 91016

RESULT 12

CCA6246 1976 bp mRNA PLN 17-AUG-1999
 LOCUS Cyanidium caldarium mRNA for glucose-6-phosphate 1-dehydrogenase.
 DEFINITION AJ006246
 ACCESSION
 VERSION AJ006246.1 GI:5734371
 KEYWORDS g6pdh gene; glucose-6-phosphate 1-dehydrogenase.
 SOURCE Cyanidium caldarium.
 ORGANISM Cyanidium caldarium.
 Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;
 Porphyridiaceae; Cyanidium.
 REFERENCE 1 (bases 1 to 1976)
 AUTHORS Wendt,U.K., Hauschild,R., Lange,C., Pietersma,M., Wenderoth,I. and

von Schaewen,A.
 Evidence for functional convergence of redox regulation in G6PDH
 isoforms of cyanobacteria and higher plants
 Plant Mol. Biol. 40 (3), 487-494 (1999)
 93364543
 REFERENCE 2 (bases 1 to 1976)
 AUTHORS von Schaewen,A.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAY-1998) von Schaewen A., Pflanzenphysiologie,
 Universitaet Osnabrueck, Barbarastr. 11, 49076 Osnabrueck, GERMANY
 FEATURES
 Location/Qualifiers
 source 1..1976
 /organism="Cyanidium caldarium"
 /db_xref="taxon:2771"
 /dev_stage="vegetative/autotrophic"
 transit_peptide 46..258
 /gene="g6pdh"
 CDS 46..1848
 /gene="g6pdh"
 /EC_number="1.1.1.49"
 /codon_start=1
 /product="glucose-6-phosphate 1-dehydrogenase"
 /protein_id="CAB52681.1"
 /db_xref="GI:5734372"

/translation="MWIPLFITFWNTNTSSSHLMKTKYTKGWTTPLYFSKVKDSFFA
 PRLASIQVQGLDNIAPQWNSLETLPIITNERPAIPROHFSHLGDTLDPSLMDTS
 VKRKQRAKEEPLCVIGAGSLDAKKTFFALFSLYVHLLPKDLVGYARRMTQ
 EEFNRNIMESLTCRVIDGPOCKWDELPKCHYMSGYDRTEDFVLQDQFLNFEQS
 PENTRVDRLYLAVPSQPVENVVHVHESGRQGRGNRIVMEKPKGKDTIYLQLRNS
 LANCISEDEIYRIDHILGKELQNLVLMFLRFANYFEPLNRDHIASIQIVFKNFGE
 GRAGIDETGIIRDIMQNLHLLQVALLGMEQVPLHAEDIRDEKVKFRLRPLKASD
 FVLQYRDQNPQSYLSEPGVMNSHTPTFAACVQVDNRNRWGVPLMKAGKALDE
 RKAELRIQFQSVPGSLFQSVYSSSHLPHNELVIIQVDEAIYMRILSKAPGTSRL
 RLNLFYTAWEDSKDIPAYERLILDVTHGKSLFIRDELEVAWNIPTPSLKEMEMA
 QDSWKPIILVGYGRGPIESDYLAAYVQNSEGD"
 46..1848
 /gene="g6pdh"
 mat_peptide 259..1845
 /gene="g6pdh"
 /EC_number="1.1.1.49"
 /product="glucose-6-phosphate 1-dehydrogenase"
 BASE COUNT 623 a 321 c 404 g 628 t
 ORIGIN

Query Match 21.6%; Score 56.8; DB 8; Length 1976;
 Best Local Similarity 50.2%; Pred. No. 0.0001;
 Matches 124; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
 QY 2 aagcacttttgatgttcgctcatgtcttgcagcagtcctcagaccagaggtgga 61
 DB 749 AAGTATTTGAAAATGTGTACATCATGTTCATGAAAGCGGAAGAACGACGAGTTGGA 808
 QY 62 atcccataatatttgagaagccatttgccttgcacttcttcccataggctgacac 121
 DB 809 ATCGCATCGTCATGAAAAACCATTTGGAAGACATTAATCTTCTATCATCGCTCGAA 868
 QY 122 aatattcttcttcaaaactttcaggaagcaaatatataagattgancatctactagaa 181
 DB 869 ATAGTCTTCGCAATTTGTATTTCTGAAGATGAATATATACCGAATTCATCATTGGA 928
 QY 182 ggaatcncagtnaaatcctcncaggtttaaagggtttcacaannagnttttgcacactt 241
 DB 929 AGGAGTTCGTAACAAATTTGATGTTTTCGCGTTTTCGCAATTTATCTATTTGAGCCTTAT 988
 QY 242 ngagng 248
 DB 989 GGAATCG 995

RESULT 13

ASU09032 463 bp mRNA INV 20-JUN-1996
 LOCUS ASU09032

Sat Nov 4 18:11:37 2000

DEFINITION Anastrepha suspensa glucose-6-phosphate dehydrogenase mRNA, partial cds.
 ACCESSION U09032
 VERSION U09032.1 GI:606621
 KEYWORDS Caribbean fruit fly.
 SOURCE Anastrepha suspensa
 ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Tephritoidea; Tephritidae; Anastrepha.

REFERENCE 1 (bases 1 to 463)
 Soto-Adames, F.N., Robertson, H.M. and Berlocher, S.H.
 Phylogenetic utility of partial DNA sequences of G6PDH at different taxonomic levels in Hexapoda with emphasis on Diptera
 Ann. Entomol. Soc. Amer. 87, 723-736 (1994)

JOURNAL 2 (bases 1 to 463)
 Soto-Adames, F.N.
 TITLE Direct Submission
 AUTHORS Soto-Adames, F.N.
 JOURNAL Submitted (19-APR-1994) Felipe N. Soto-Adames, Entomology, University of Illinois, 505 S. Goodwin Ave., Urbana, IL 61801, USA

FEATURES
 Source 1..463
 Location/Qualifiers
 /organism="Anastrepha suspensa"
 /db_xref="taxon:28587"
 /note="partial sequence from PCR product"
 <1..>463
 /codon_start=2
 /product="glucose-6-phosphate dehydrogenase"
 /protein_id="AAB02776.1"
 /db_xref="GI:606621"
 /translation="LWLFQRLNLLPKPTKFGYARSKLTVDLRAKCHQVMKVPQHEQ
 KYDEFQWYNYVAGGVDQSDFDALRDLQRLDCDCNNRFLYALPSPVQATV
 NKIDICQSECGWNNVIEKPPGDDASKKLSHLSLFEEBQIYRIDHVLG"

BASE COUNT 119 a 107 c 113 g 124 t
 ORIGIN

Query Match 21.5%; Score 56.6; DB 30; Length 463;
 Best Local Similarity 61.8%; Pred. No. 0.00011;
 Matches 89; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 37 agtgcacagaccaggaagatcgacataatttgagaagccatttgcttgat 96
 Db 320 ATTTGTCATCGAATGCGGTGGAATCGGTGATCGAATCGAATCGGCGGTGAC 379

QY 97 gcaatttcctccataggtgcacacataatttttcaaaccttcaggaagcaata 156
 Db 380 GATGCGCTCTTCAAGAACTGAGCGACCATCTGCTTCACTCTTCCACGAGGCAATC 439

QY 157 tatagaatgancatctacttagga 180
 Db 440 TATCGTATGATCATCTCTGGA 463

RESULT 14
 AF269359 3015 bp DNA BCT 03-AUG-2000
 LOCUS Staphylococcus epidermidis strain SR1 clone step.1002e12 genomic sequence.

ACCESSION AF269359
 VERSION AF269359.1 GI:9664680
 KEYWORDS Staphylococcus epidermidis.
 SOURCE Staphylococcus epidermidis.
 ORGANISM Staphylococcus epidermidis.
 Bacillus/Firmicutes; Bacillus/Clostridium group; Staphylococcus.

REFERENCE 1 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

JOURNAL 2 (bases 1 to 3015)
 Kimmerly, W.J., Taylor, J., David, Nelsen, A.J., Godlevski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P.J.
 Transposon-mediated sequencing of the Staphylococcus epidermidis

FEATURES Drive, Research Triangle Park, North Carolina 27709-3398, USA
Location/Qualifiers

source 1..3081
/organism="Staphylococcus epidermidis"
/strain="SR1"
/db_xref="taxon:1282"
/clone="step.1010e05"

BASE COUNT 963 a 528 c 419 g 1171 t
ORIGIN

Query Match 20.5%; Score 54; DB 1; Length 3081;
Best Local Similarity 54.5%; Pred. No. 0.00053;
Matches 114; Conservative 0; Mismatches 93; Indels 2; Gaps 1;
Qy 45 qaccagaaggatgcataatatttgagaagccatttgctttgatgcacttc 104
Db 2346 GAACGTAAAGGGGTGATGCGCAATATATAGAAAAACCATTTGGTGATGATTAAAAATC 2287
Qy 105 ttccataggtgcacacatatctcttccaaactttcaggaaaaagcaaatatataaat 164
Db 2286 T--GCAAAAAAATTAAATAATCAATAAGAAAGTCTTTTAAGAGAAAAAATATTTGGAAT 2229
Qy 165 tgancatctactaggaaggaatcncagtnaaatccctncaggtttaagggttttcaaan 224
Db 2228 TGATCAGTATTAGGTAAAGAAATGATTCAAAATATTGAACGCTACGATTTTGAACAC 2169
Qy 225 agntttgagccaccttngagngnaent 253
Db 2168 TATATTTGAACCCACTTTTGGAAATAATAAGT 2140

Search completed: November 4, 2000, 13:30:39
Job time: 16888 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 13:40:19 ; Search time 320.8 Seconds
(without alignments)
307.978 Million cell updates/sec

Title: US-09-300-482-4

Perfect score: 263

Sequence: 1 gaagcacttttgatgtgc.....agnngnacntnnnnnganna 263

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_36.*
1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.2	22.1	32768	20	X13065
2	51.4	19.5	16592	18	V74364
3	51	19.4	2750	14	Q38608
C 4	50.6	19.2	2494	21	A26852
5	49.6	18.9	1488	19	V32895
C 6	48	18.3	2764	19	V65241
7	48	18.3	13121	19	V52259
C 8	44.8	17.0	1488	19	V56039
9	44.8	17.0	10820	20	X20527
10	42.2	16.0	1482	11	Q05236
11	39.8	15.1	554	20	X21060
12	39.6	15.1	2260	18	T88030
					Enterococcus faeca
					Staphylococcus aur
					G6PD coding sequen
					Essential Staphylo
					Streptococcus pneu
					DNA encoding a S.
					Streptococcus pneu
					Thermotable gluco
					Polynucleotide seq
					Glucose-6-phosphat
					Polynucleotide seq
					Brevibacterium fla

13	39	14.8	780	18	T68126
14	37.4	14.2	1461	16	T30576
15	37.2	14.1	1455	16	T30577
16	37.2	14.1	1467	16	T30575
17	37.2	14.1	1696	13	Q20959
C 18	34.4	13.1	1981	20	X84323
19	33	12.5	1380	21	Z34730
20	33	12.5	1488	21	X34648
21	33	12.5	2042	20	X33972
22	33	12.5	2266	21	Z45852
C 23	33	12.5	9199	20	X84322
24	32.6	12.4	1636	21	Z65255
25	32.4	12.3	600	20	V86185
C 26	32	12.2	87350	18	X83003
C 27	31.6	12.0	1230	10	N90223
28	31.4	11.9	1230025	20	X91990
29	30.4	11.6	1335	20	Z52939
30	30.4	11.6	237326	19	V57903
C 31	30.2	11.5	65921	21	Z89046
32	30	11.4	1380	21	Z34650
33	29.8	11.3	1458	16	T30574
34	29.8	11.3	1963	13	Q24526
35	29.8	11.3	11901	20	X02998
36	29.6	11.3	1484	20	X89755
37	29.6	11.3	1517	14	Q47166
38	29.6	11.3	1517	14	Q47167
39	29.6	11.3	1517	19	V23923
40	29.6	11.3	1517	19	V24099
41	29.6	11.3	1517	19	V25817
42	29.6	11.3	1517	19	V25818
43	29.6	11.3	1517	19	V09855
44	29.6	11.3	1517	19	V09856
45	29.6	11.3	1517	19	V12162

ALIGNMENTS

RESULT 1

X13065
ID X13065 standard; DNA; 32768 BP.

XX X13065;

XX 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:128.

XX Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

PN W09850555-A2.

XX 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.

H. pylori cytoplas
Leuconostoc lactis
Leuconostoc dextra
Leuconostoc citreu
L.dextranicus gluc
Stealth virus nucl
Scarface 3 gene co
Human growth facto
Human TIE ligand N
Human liver angio
Stealth virus nucl
Human secreted pro
EST clone J635 H
Human WRN genomic
Malaria-specific D
Nucleotide sequenc
Human prostate tum
Hereditary haemoch
Human nibrin DNA
Human growth facto
Leuconostoc mesent
Glc6PD genome. Le
Human IL-1ra BAC c
Probe sequence for
BPV1 L1 ORF. Bovi
HPV1 L1 ORF. Huma
HPV16 L1 gene #1.
HPV16 L1 gene #2.
Human papillomavir
Human papillomavir
HPV16 L1 capsid pr
HPV16 recombinant
Human papillomavir

[illegible]

CC of the above nucleic acids can be used for the recombinant expression of
 CC the protein sequences. The invention also provides a DNA chip having
 CC arrayed on it at least 15 base pair fragment of any one or more of these
 CC DNA sequences. The DNA chip can be used methods for evaluating gene
 CC expression in *S. pneumoniae* and for identifying virulence genes in
 CC *S. pneumoniae*. Antibodies that selectively bind to the above proteins or
 CC peptide fragments can be used to treat *S. pneumoniae* infection. The
 CC antibodies can also be used to detect *S. pneumoniae* cells.
 XX
 SQ Sequence 2764 BP; 829 A; 515 C; 646 G; 774 T; 0 other;

Query Match 18.3%; Score 48; DB 19; Length 2764;
 Best Local Similarity 51.3%; Pred. No. 4.6e-05;
 Matches 99; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
 QY 52 aaggatggaatgcataatattgagagcatttggttgatgcacttttcccat 111
 Db 2743 AAAGGTTTTGAGCGCTTGATCGTTGAAAACCAATTTGTACAGATTACGCAACTGCAAC 2684
 QY 112 aggtgacacaaatattcttcaaaactttcaggaagcaaatatataagaattgancat 171
 Db 2683 AGTTGATGACGAACTCTAGCAACATTTGACGAAGAACAAATTTCCGTATCGACCAT 2624
 QY 172 ctactaggaagaatcncagtnaaatcctncaggttttaagggtttcaaanngntttt 231
 Db 2623 TATCTTGTGAAGGAATGATCCAAAGCATCTTTGCAGTTCGCTTTCGAAACTTGATTTT 2564
 QY 232 gaggcacacttnga 244
 Db 2563 GAAACGTTTGGG 2551

RESULT 7
 V52259
 ID V52259 standard; DNA; 13121 BP.
 XX
 AC V52259;
 XX
 DT 23-OCT-1998 (first entry)
 XX
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:126.
 XX
 KW Streptococcus pneumoniae; *S. pneumoniae*; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN W09818931-A2.
 XX
 PD 07-MAY-1998.
 XX
 PF 30-OCT-1997; 97WO-US19588.
 XX
 PR 31-OCT-1996; 96US-0029960.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillion PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 DR WPI; 1998-272225/24.
 XX
 PT Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 XX
 PS Claim 1; Page 869-876; 1409pp; English.
 XX
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
 CC on it, or a representative fragment or a sequence at least 95% identical

CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
 CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
 CC pneumoniae. The present invention also describes an isolated nucleic acid
 CC molecule encoding a homologue of any of the fragments of the *S. pneumoniae*
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
 CC by a process comprising: (a) screening a genomic DNA library using as a
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
 CC to 391, identifying members of the library which contain sequences
 CC that hybridize to the target sequence and isolating the nucleic acid
 CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
 CC from an organism, amplifying nucleic acid molecules whose nucleotide
 CC sequence is homologous to amplification primers derived from the
 CC fragment of the *S. pneumoniae* genome to prime the amplification and
 CC isolating the amplified sequences. The computer readable medium can be
 CC used in a computer-based system for identifying fragments of the
 CC *S. pneumoniae* genome of commercial importance, or expression/modulating
 CC fragments of the *S. pneumoniae* genome. Products from the present
 CC invention can be used in diagnosis kits and assays, and pharmaceutical
 CC compositions and vaccines for *S. pneumoniae*.

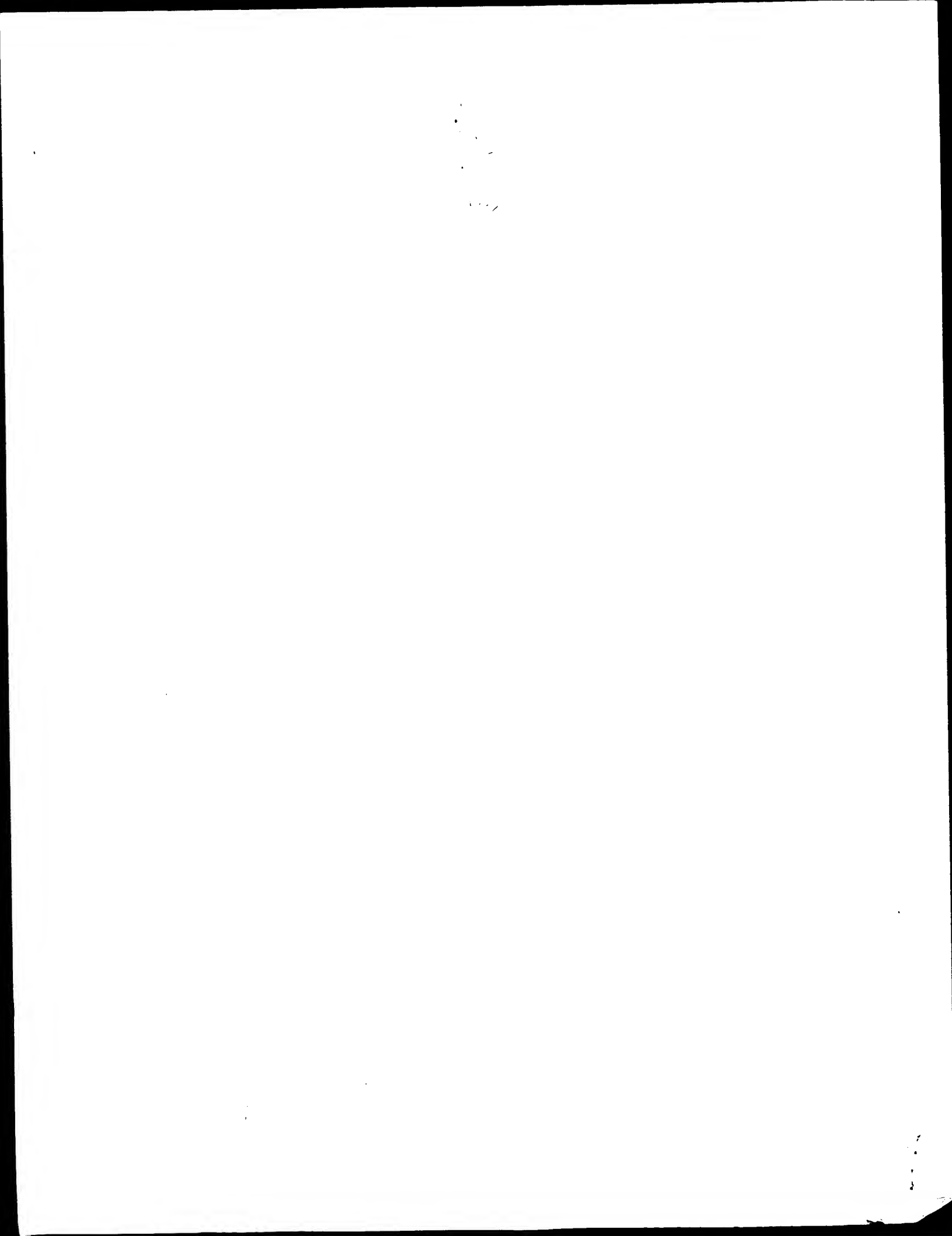
XX Sequence 13121 BP; 3698 A; 3021 C; 2439 G; 3963 T; 0 other;

Query Match 18.3%; Score 48; DB 19; Length 13121;
 Best Local Similarity 48.1%; Pred. No. 7.5e-05;
 Matches 102; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
 QY 52 aaggatggaatgcataatattgagagcatttggttgatgcacttttcccat 111
 Db 3492 aaagggtttgagcgttgatcgcttgaaaaccatttggtacagattacgcaactgcaagc 3551
 QY 112 aggtgacacaaatattcttcaaaactttcaggaagcaaatatataagaattgancat 171
 Db 3552 aagttgaatgacgaactcctagcaacatttgacgaagaacaaatttcgtagcgaccat 3611
 QY 172 ctactaggaagaatcncagtnaaatcctncaggttttaagggtttcaaanngntttt 231
 Db 3612 tatctgtgtaagaaatgatccaaagcatcttgcggttcgctttgcaaacattgatttt 3671
 QY 232 gaggcacacttngagngnacntnnnnnganna 263
 Db 3672 gaaacgtttggaacaaggattttatcgacaa 3703

RESULT 8
 V56039
 ID V56039 standard; DNA; 1488 BP.
 XX
 AC V56039;
 XX
 DT 09-DEC-1998 (first entry)
 XX
 DE Thermostable glucose-6-phosphate dehydrogenase encoding DNA.
 XX
 DE Glucose-6-phosphate dehydrogenase; thermostable; heat-resistant;
 KW enzyme; recombinant; high stability; ds.
 KW
 KW
 XX
 OS *Pedococcus pentosaceus*.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1488
 FT /*tag= a
 FT /product= "glucose-6-phosphate dehydrogenase"
 XX
 PN JP10225293-A.
 XX
 PD 25-AUG-1998.
 XX
 PF 13-FEB-1997; 97JP-0029345.
 XX
 PR 13-FEB-1997; 97JP-0029345.
 XX
 PA (KIRK) KIRKMAN CORP.

```
XX WPI; 1998-513903/44.
DR P-PSDB; W80567.
XX Heat-resistant glucose-6-phosphate dehydrogenase gene - useful for
PT efficiently preparing heat-resistant glucose-6-phosphate
PT dehydrogenase of high stability
XX
XX Claim 1; Pages 5-6; 7pp; Japanese.
XX
XX This DNA encodes a heat-resistant glucose-6-phosphate dehydrogenase. A
CC transformant or a transductant containing a recombinant DNA in which the
CC heat-resistant (thermostable) glucose-6-phosphate dehydrogenase gene is
CC inserted to a vector DNA, can be cultured to produce the enzyme
CC recombinantly. The method can prepare heat-resistant glucose-6-phosphate
CC dehydrogenase of high stability efficiently.
XX
XX Sequence 1488 BP; 496 A; 249 C; 308 G; 435 T; 0 other;
SQ
Query Match 17.0%; Score 44.8; DB 19; Length 1488;
Best Local Similarity 47.2%; Pred. No. 0.00033;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
QY 52 aaggatggaatcgatcataatattgagagccatttggttgatgcacttttcccat 111
DB || || || || || || || || || || || || || || || || || || || ||
QY 112 aggtgacacatactcttttcaaaccttcaggaaagcaaatatagaaattgancat 171
DB || || || || || || || || || || || || || || || || || || || ||
QY 490 gaatgaatgattcaatttcgggttccttaagaagaacagatttttagaattgatcat 549
DB || || || || || || || || || || || || || || || || || || || ||
QY 172 ctactaggaagaatcnaagtnaaatcctncaggttttaagggttttcaaaannnagntttt 231
DB || || || || || || || || || || || || || || || || || || || ||
QY 550 tatttagttaaggaaacccatccagtcacattgctgattacggtttgggaacgcatgttc 609
QY 232 gagccaccttngnngnacntnnnnnganna 263
DB || || || || || || || || || || || || || || || || || || || ||
QY 610 aactcgtttggaataagaagaacatttgataa 641
RESULT 9
X20527/C
ID X20527 standard; DNA; 10820 BP.
XX
XX X20527;
XX
XX 05-MAY-1999 (first entry)
XX Polynucleotide sequence from the genome of Treponema pallidum.
XX Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX enzyme production; ds.
XX
XX Treponema pallidum.
XX
XX W09859034-A2.
XX
XX 30-DEC-1998.
XX
XX 23-JUN-1998; 98WO-US13041.
XX
XX 24-JUN-1997; 97US-0050667.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fraser CM;
XX
XX WPI; 1999-081273/07.
XX
XX New isolated Treponema pallidum nucleic acids - used to develop
XX products for the detection, diagnosis, characterisation, prevention
XX and therapy of T. pallidum infections, particularly syphilis
```

```
XX Claim 1; Page 343-349; 1150pp; English.
XX
XX X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
XX Sequence 10820 BP; 2418 A; 2982 C; 2962 G; 2450 T; 8 other;
SQ
Query Match 17.0%; Score 44.8; DB 20; Length 10820;
Best Local Similarity 48.0%; Pred. No. 0.00062;
Matches 94; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 68 taatattgagaagccatttggttgatgcacttttcccatagggtgacacaaatc 127
DB || || || || || || || || || || || || || || || || || || || ||
QY 128 ttctttcaacttccaggaagcaaatatagaaattgancatctactaggaaggaatc 187
DB || || || || || || || || || || || || || || || || || || || ||
QY 6607 TCGTGCCCACTTTTCAGGAAACCAACCTATCGCATGATCATTCTCTGGTAAGGAAA 6548
QY 188 ncagtnaaatcctncaggttttaagggttttcaaaannnagnttttgagccaccttngann 247
DB || || || || || || || || || || || || || || || || || || || ||
QY 6547 CGGTCCAAACATCCTGGTCACCTCGCTTTGCCAATCCCTTTTCGAGCCGACATGGAAC 6488
QY 248 gnaactnnnnnganna 263
DB || || || || || || || || || || || || || || || || || || || ||
QY 6487 GGACCCATATCGATTA 6472
RESULT 10
Q05236
ID Q05236 standard; DNA; 1482 BP.
XX
XX Q05236;
XX
XX 12-NOV-1990 (first entry)
XX
XX Glucose-6-phosphate dehydrogenase gene.
XX
XX Glucose-6-phosphate dehydrogenase; creatine kinase; ss.
XX
XX Bacillus.
XX
XX Key Location/Qualifiers
XX CDS 1..1482
XX FT /*tag= a
XX
XX DE3942129-A.
XX
XX 05-JUL-1990.
XX
XX 20-DEC-1989; 89DE-3942129.
XX
XX 29-DEC-1988; 88JP-0334629.
XX
XX (TOXN ) TOYO JOZO KK.
XX
XX Sagai H, Hattori K, Takahashi M;
XX
XX WPI; 1990-211190/28.
XX
XX P-PSDB; R05798.
XX
XX New glucose-6-phosphate dehydrogenase gene from Bacillus species -
XX used for expressing stable, very pure enzyme, useful in assay of
XX creatine kinase.
XX
XX Disclosure; ; p; German.
XX
```

Best Local Similarity 44.7%; Pred. No. 0.0059;
Matches 98; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 45 gaccagaaggatggaatcgataataatttgagaagccatttgcttgatgcacttc 104
Db 411 GCCCAACACTGGTTACACCGTTTGATGATCGAAAAGCCATTGGGACATCATACGAAC 470
QY 105 ttccatagctgacacatactcttcttcaaacctttcaggaagcaaaatataagaat 164
Db 471 AGCTGAAAGTTGCAAAACGAATTGGAAAACGCCCTTTGATGATGACCAATTTGTTCCGTAT 530
QY 165 tgancatctactaggaagaatcncagtnaaaatcctncaggtttaagggtttcaaan 224
Db 531 TGACCACACTACCTTGGTAAGGAATGGTCCAAAATATTGCGGCTTTGGCTTTGTAACCC 590
QY 225 agntttgagccaccttngagngnacntnnnnnganna 263
Db 591 AATCTTTGATGACGCTGGAAACAGGACTACATCAAGAA 629

RESULT 4

US-08-445-464C-5

; Sequence 5, Application US/08445464C

; Patent No. 6090567

; GENERAL INFORMATION:

; APPLICANT: Jakobovits, Edward B.

; APPLICANT: Silen, Joy L.

; APPLICANT: Levy, Mark J.

; APPLICANT: Goodman, Thomas C.

; APPLICANT: Becker, Martin

; APPLICANT: Ullman, Edwin F.

; APPLICANT: Caldwell, Robert M.

; APPLICANT: Bott, Richard R.

; APPLICANT: Barnett, Christopher C.

; TITLE OF INVENTION: Homogeneous Immunoassays Using Mutant

; Glucose-6-Phosphate Dehydrogenases

; NUMBER OF SEQUENCES: 124

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dade Behring Inc.

; STREET: 1717 Deerfield Road

; CITY: Deerfield

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60015

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/445,464C

; FILING DATE: 22-MAY-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/044,857

; FILING DATE: 08-APR-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Ruzsala, Lois K.

; REGISTRATION NUMBER: 39,074

; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1

; TELEPHONE: (847) 267-5364

; TELEFAX: (847) 267-5376

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1461 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

US-08-445-464C-5

Query Match 14.2%; Score 37.4; DB 5; Length 1461;
Best Local Similarity 44.7%; Pred. No. 0.0059;
Matches 98; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 45 gaccagaaggatggaatcgataataatttgagaagccatttgcttgatgcacttc 104
Db 411 GCCCAACACTGGTTACACCGTTTGATGATCGAAAAGCCATTGGGACATCATACGAAC 470
QY 105 ttccatagctgacacatactcttcttcaaacctttcaggaagcaaaatataagaat 164
Db 471 AGCTGAAAGTTGCAAAACGAATTGGAAAACGCCCTTTGATGATGACCAATTTGTTCCGTAT 530
QY 165 tgancatctactaggaagaatcncagtnaaaatcctncaggtttaagggtttcaaan 224
Db 531 TGACCACACTACCTTGGTAAGGAATGGTCCAAAATATTGCGGCTTTGGCTTTGTAACCC 590
QY 225 agntttgagccaccttngagngnacntnnnnnganna 263
Db 591 AATCTTTGATGACGCTGGAAACAGGACTACATCAAGAA 629

RESULT 5

PCT-US94-03437-5

; Sequence 5, Application PC/TUS9403437

; GENERAL INFORMATION:

; APPLICANT:

; APPLICANT:

; TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT

; Glucose-6-PHOSPHATE DEHYDROGENASES

; NUMBER OF SEQUENCES: 124

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/03437

; FILING DATE:

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1461 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA genomic

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Leuconostoc lactis

; STRAIN: NCDO 546

PCT-US94-03437-5

Query Match 14.2%; Score 37.4; DB 6; Length 1461;
Best Local Similarity 44.7%; Pred. No. 0.0059;
Matches 98; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 45 gaccagaaggatggaatcgataataatttgagaagccatttgcttgatgcacttc 104
Db 411 GCCCAACACTGGTTACACCGTTTGATGATCGAAAAGCCATTGGGACATCATACGAAC 470
QY 105 ttccatagctgacacatactcttcttcaaacctttcaggaagcaaaatataagaat 164
Db 471 AGCTGAAAGTTGCAAAACGAATTGGAAAACGCCCTTTGATGATGACCAATTTGTTCCGTAT 530
QY 165 tgancatctactaggaagaatcncagtnaaaatcctncaggtttaagggtttcaaan 224
Db 531 TGACCACACTACCTTGGTAAGGAATGGTCCAAAATATTGCGGCTTTGGCTTTGTAACCC 590
QY 225 agntttgagccaccttngagngnacntnnnnnganna 263

Db 591 AATCTTTGATGCGCTGCTGAACAGGACTATCAAGAA 629

```

RESULT 6
PCT-US94-03437-7
; Sequence 7, Application PC/TUS9403437
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES
; NUMBER OF SEQUENCES: 124
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03437
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA genomic
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Leuconostoc dextranicum
; STRAIN: ATCC 19255
; PCT-US94-03437-7

```

```

Query Match 14.1%; Score 37.2; DB 6; Length 1455;
Best Local Similarity 44.9%; Pred. No. 0.0068;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 50 agaaggatggaatcgacataattttgagaagccatttgcttgcacatttcttccc 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 AGACTGGTACATCTTTGATGATTTGAAAGCGCTTTGGTACATCATACGCCGCGC 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 110 ataggctgacacaatatcttcttcaaaccttcaggaaagcaaatatataagaattganc 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 AAGAATTGCAAAAGTATTTGGAAATGCAATTTGATGATGACCAACTGTCCTGTTGACC 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 170 atctactagaaggaatcncagtnaaatcctncagtttaagggtttcaaaannnagntt 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 ACTATCTTGGAAAGAAATGGTACAAATATTGACGACATTACGTTTGTGTAACCCCAATCT 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 230 ttgagccacctngagnnagncnnnnnnnanna 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 TTGATGCGCTTGGAAATGAGGACTATATCAAAAA 626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
PCT-US94-03437-3
; Sequence 3, Application PC/TUS9403437
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HOMOGENEOUS IMMUNOASSAYS USING MUTANT
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASES
; NUMBER OF SEQUENCES: 124
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/03437
; FILING DATE:

```

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1467 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA genomic
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Leuconostoc citreum
; STRAIN: NCIMB 3351
; PCT-US94-03437-3

Query Match 14.1%; Score 37.2; DB 6; Length 1467;
Best Local Similarity 44.9%; Pred. No. 0.0068;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 50 agaaggatggaatcgacataattttgagaagccatttgcttgcacatttcttccc 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 413 ATAGTGGTTACAAACCGTTTTCATGATTTGAAAGCGCTTTGGTACATCATACGCCGCGC 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 110 ataggctgacacaatatcttcttcaaaccttcaggaaagcaaatatataagaattganc 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 473 AAGAGCTACAAAAAGACTTAGAAAAAGCGCTTTTGACGATAATCAATTTATCCCGTATTGATC 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 170 atctactagaaggaatcncagtnaaatcctncagtttaagggtttcaaaannnagntt 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 533 ATTATCTTGGTAAAGAAATGGTCCAAATATTGCTGCCCTTCGTTTTTGGTAACCCCATCT 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 230 ttgagccacctngagnnagncnnnnnnnanna 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 593 TTGATGCGCTTGGAAACAAAGATTACATTAATAAAA 626
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
US-08-445-463B-3
; Sequence 3, Application US/08445463B
; Patent No. 6033890
; GENERAL INFORMATION:
; APPLICANT: Jakobovits, Edward B.
; APPLICANT: Silen, Joy L.
; APPLICANT: Levy, Mark J.
; APPLICANT: Goodman, Thomas C.
; APPLICANT: Becker, Martin
; APPLICANT: Ullman, Edwin F.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Bott, Richard R.
; APPLICANT: Barnett, Christopher C.
; TITLE OF INVENTION: Homogenous Immunoassays Using Mutant
; TITLE OF INVENTION: Glucose-6-Phosphate Dehydrogenases
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dade Behring Inc.
; STREET: 1717 Deerfield Road
; CITY: Deerfield
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60015
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,463B
; FILING DATE: 22-MAY-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/044,857
; FILING DATE: 08-APR-1993
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Ruzsala, Lois K.
; REGISTRATION NUMBER: 39,074
; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 267-5364
; TELEFAX: (847) 267-5376
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-445-463B-3

```

```

Query Match      14.1%   Score 37.2; DB 5; Length 1571;
Best Local Similarity 44.9%; Pred. No. 0.007;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 50 agaagggatggaatcgcaataattttgagaagccatttggcttggatgcacatttcttccc 109
Db 501 ATAGTGGTTACACCGTTTGATGATTGAAAGCCCTTTGGTACATCATAGCCCACTGCCG 560

QY 110 ataggctgacacaatatcttcttcaaaacttcaggaaagcaaatatagaaatganc 169
Db 561 AAGAGCTACAAAAGACTTAGAAAAGCGCTTTGACGATAATCAATATTCGGTATGATC 620

QY 170 atctactaggaagaaatcncagtnaaatcctncaggtttaagggtttcaaaannnagntt 229
Db 621 ATTATCTTGGTAAAGAAATGTCGCAAAATATTGCTGCCCTTCGTTTGGTAAACCCCATCT 680

QY 230 ttgagccaccttgagngnagcnnnnnnnanna 263
Db 681 TTGATGCCGCTTGGACAAAGATTACATTAAAAA 714

```

```

RESULT 9
US-08-445-464C-3
; Sequence 3, Application US/08445464C
; Patent No. 6090567
; GENERAL INFORMATION:
; APPLICANT: Jakobovits, Edward B.
; APPLICANT: Sillen, Joy L.
; APPLICANT: Levy, Mark J.
; APPLICANT: Goodman, Thomas C.
; APPLICANT: Becker, Martin
; APPLICANT: Ullman, Edwin F.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Bott, Richard R.
; APPLICANT: Barnett, Christopher C.
; TITLE OF INVENTION: Homogenous Immunoassays Using Mutant
; TITLE OF INVENTION: Glucose-6-Phosphate Dehydrogenases
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dade Behring Inc.
; STREET: 1717 Deerfield Road
; CITY: Deerfield
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60015
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445.464C
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/044,857
; FILING DATE: 08-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ruzsala, Lois K.
; REGISTRATION NUMBER: 39,074
; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 267-5364
; TELEFAX: (847) 267-5376
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1571 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-445-464C-3

```

```

Query Match      14.1%   Score 37.2; DB 5; Length 1571;
Best Local Similarity 44.9%; Pred. No. 0.007;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 50 agaagggatggaatcgcaataattttgagaagccatttggcttggatgcacatttcttccc 109
Db 501 ATAGTGGTTACACCGTTTGATGATTGAAAGCCCTTTGGTACATCATAGCCCACTGCCG 560

QY 110 ataggctgacacaatatcttcttcaaaacttcaggaaagcaaatatagaaatganc 169
Db 561 AAGAGCTACAAAAGACTTAGAAAAGCGCTTTGACGATAATCAATATTCGGTATGATC 620

QY 170 atctactaggaagaaatcncagtnaaatcctncaggtttaagggtttcaaaannnagntt 229
Db 621 ATTATCTTGGTAAAGAAATGTCGCAAAATATTGCTGCCCTTCGTTTGGTAAACCCCATCT 680

QY 230 ttgagccaccttgagngnagcnnnnnnnanna 263
Db 681 TTGATGCCGCTTGGACAAAGATTACATTAAAAA 714

```

```

RESULT 10
US-07-737-071A-1
; Sequence 1, Application US/07737071A
; Patent No. 529286
; GENERAL INFORMATION:
; APPLICANT: JARSCH, Michael
; APPLICANT: LANG, Gunter
; TITLE OF INVENTION: CLONING AND OVEREXPRESSION OF
; TITLE OF INVENTION: GLUCOSE-6-PHOSPHATE DEHYDROGENASE FROM LEUCONOSTOC
; TITLE OF INVENTION: DEXTRANICUS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Nikaido, Marmelstein Kubovcik &
; STREET: 1725 K Street, N.W., Suite 1000
; CITY: Washington D.C.
; COUNTRY: United States of America
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/737,071A
; FILING DATE: 19910730
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 4024158.9
; FILING DATE: 30-JUL-1990
; ATTORNEY/AGENT INFORMATION:

```


; FILING DATE: 22-MAY-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/044,857
; FILING DATE: 08-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ruzsala, Lois K.
; REGISTRATION NUMBER: 39,074
; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 267-5364
; TELEFAX: (847) 267-5376
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1696 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-445-463B-7

Query Match 14.1%; Score 37.2; DB 5; Length 1696;
Best Local Similarity 44.9%; Pred. No. 0.0072;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 50 agaagggatgaacgcataatatttgagaagccatttgcttgatgcactttcttccc 109
DB 538 AGACTGGCTACATCGTTTGATGATTGAAAGCCCTTTTGGTACATCATACGCCACCGCAG 597
QY 110 ataggctgacacaatattcttcttcaaaccttcaggaaagcaaatatagaaattganc 169
DB 598 AAGAATTGCAAGTGATTGGAAATGCATTGTGATGACCAACTGTCCTGATTGACC 657
QY 170 atctactaggaagaatcncagtnaaatcctncaggttttaaggggtttcaaanngntt 229
DB 658 ACTATCTTGGAAAGAAATGGTACAAAATATTGCAGCATTACGTTTGGTAACCCAATCT 717
QY 230 ttgagccaccttngagngnacntnnnnnganna 263
DB 718 TTGATGCCGCTTGGAAATAGGACTATATCAAAAA 751

RESULT 13
US-08-445-464C-7
; Sequence 7, Application US/08445464C
; Patent No. 6090567
; GENERAL INFORMATION:
; APPLICANT: Jakobovits, Edward B.
; APPLICANT: Silen, Joy L.
; APPLICANT: Levy, Mark J.
; APPLICANT: Goodman, Thomas C.
; APPLICANT: Becker, Martin
; APPLICANT: Ullman, Edwin F.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Bott, Richard R.
; APPLICANT: Barnett, Christopher C.
; TITLE OF INVENTION: Homogenous Immunoassays Using Mutant
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dade Behring Inc.
; STREET: 1717 Deerfield Road
; CITY: Deerfield
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60015
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,464C
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/044,857
; FILING DATE: 08-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Ruzsala, Lois K.
; REGISTRATION NUMBER: 39,074
; REFERENCE/DOCKET NUMBER: BEH-7261 DIV 1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 267-5364
; TELEFAX: (847) 267-5376
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1696 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-445-464C-7

Query Match 14.1%; Score 37.2; DB 5; Length 1696;
Best Local Similarity 44.9%; Pred. No. 0.0072;
Matches 96; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 50 agaagggatgaacgcataatatttgagaagccatttgcttgatgcactttcttccc 109
DB 538 AGACTGGCTACATCGTTTGATGATTGAAAGCCCTTTTGGTACATCATACGCCACCGCAG 597
QY 110 ataggctgacacaatattcttcttcaaaccttcaggaaagcaaatatagaaattganc 169
DB 598 AAGAATTGCAAGTGATTGGAAATGCATTGTGATGACCAACTGTCCTGATTGACC 657
QY 170 atctactaggaagaatcncagtnaaatcctncaggttttaaggggtttcaaanngntt 229
DB 658 ACTATCTTGGAAAGAAATGGTACAAAATATTGCAGCATTACGTTTGGTAACCCAATCT 717
QY 230 ttgagccaccttngagngnacntnnnnnganna 263
DB 718 TTGATGCCGCTTGGAAATAGGACTATATCAAAAA 751

RESULT 14
US-08-933-821-16
; Sequence 16, Application US/08933821
; Patent No. 5972338
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,821
; FILING DATE:
; CLASSIFICATION: 536

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 11:51:59 ; Search time 4352.3 Seconds
(without alignments)
373.614 Million cell updates/sec

Title: US-09-300-482-4
Perfect score: 263

Sequence: 1 gaagcacttttgatgtgc.....agngnacntnnnnnganna 263

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estba.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estcom.*
68: em_estov1.*
69: em_estov2.*
70: em_estpl1.*
71: em_estpl2.*
72: em_estpl3.*
73: em_estpl4.*
74: em_estpl5.*
75: em_estrol.*
76: em_estrol2.*
77: em_estrol3.*
78: em_estrol4.*
79: em_estrol5.*
80: em_estrol6.*
81: em_estrol7.*
82: em_estrol8.*
83: em_estrol9.*
84: em_estrol10.*
85: em_estrol11.*
86: em_estrol12.*
87: em_estrol13.*
88: gb_gss1.*
89: gb_gss2.*
90: gb_gss3.*
91: gb_gss4.*
92: em_gss1.*
93: em_gss2.*
94: em_gss3.*
95: em_gss4.*
96: gb_gss5.*
97: gb_gss6.*
98: gb_gss7.*
99: gb_gss8.*
100: gb_gss9.*
101: em_gss5.*
102: em_gss6.*
103: em_gss7.*
104: em_gss8.*
105: em_gss9.*
106: em_gss10.*
107: em_gss11.*
108: gb_gss10.*
109: gb_gss11.*
110: em_gss12.*
111: gb_gss12.*
112: gb_gss13.*
113: gb_gss14.*
114: gb_gss15.*
115: gb_gss16.*
116: gb_gss17.*

VERSION AW335839.1 GI:6832460

EST.

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS

Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Laegreid, W.W., and Keele, J.W.

TITLE

Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle

JOURNAL

COMMENT

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904 e. Vector identified by cross_match with the -minscore 20

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCAGCAG

Plate: 15 row: B column: 17

Seq primer: ATTGAGTCACATATAG.

Location/Qualifiers

1..295

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 1BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

69 a 84 c 81 g 61 t

BASE COUNT

ORIGIN

Query Match 19.2%; Score 50.6; DB 21; Length 295;

Best Local Similarity 51.0%; Pred. No. 0.00015;

Matches 104; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 45 gaccagagaggaatgcataattttgagaagccatttggctttgatgcatttc 104

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 6 GAGCCAGACAGCTGGACCCATCATCGTGGAGAACCGGTTGGGAGGACCTGCACAG 65

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 105 ttccataggtgcacacatatcttctcaaacctttcaggaaagcaaatatagaat 164

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 66 CTCCAACAGCTGTCCAACACATCGCTCCCTGTTCACGAGACCATCTACCGCAT 125

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 165 tgancatctactaggaaggaatcncagtnaaatcctncaggtttaaggtttcaaan 224

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 126 TGACACTACTCGGGGAAAGAGATGGTCCAGAACCTCATGTGCTGAGTTTGCAACAG 185

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 225 agntttgagccacttngagngn 248

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 186 GATCTCGGGCCATTTGGACCG 209

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3

BE526947

LOCUS

DEFINITION

BE526947

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

M66J03STM Arabidopsis developing seed Arabidopsis thaliana cDNA

clone 600035613R1 5', mRNA sequence.

260 bp mRNA EST 09-AUG-2000

GI:9784848

thale cress.

Arabidopsis thaliana

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 260)

White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de

Ilarduya, O., Jaworski, J.G., Ohlrogge, J., and Benning, C.

A new set of Arabidopsis ESTs from developing seeds: the metabolic

pathway from carbohydrates to seed oil

Plant Physiol. (2000) In press

Contact: Benning, C

Dept. of Biochemistry & Molecular Biology

Michigan State University

224 Biochemistry, Michigan State University, East Lansing, MI 48824

, USA

Tel: 517 355 1609

Fax: 517 353 9334

Email: benning@msu.edu

Clones were originally prepared at Michigan State University.

Arabidopsis Biological Resource Center, The Ohio State University,

309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210

USA, FAX: 6142920603 TEL: 6142929371.

FEATURES

Location/Qualifiers

1..260

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone_lib="600035613R1"

/tissue_type="seed"

/dev_stage="5-13 days after flowering"

/lab_host="E.coli"

/note="Organ: Developing seed; Vector: pBluescript SK-;

Site_1: EcoRI; Site_2: XhoII"

74 a 49 c 63 g 74 t

BASE COUNT

ORIGIN

Query Match 18.6%; Score 48.8; DB 36; Length 260;

Best Local Similarity 50.9%; Pred. No. 0.00047;

Matches 110; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 6 acttttgatgttcgtctcatgtcttgcagcagtgctcagagccagagggatggaatcg 65

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 44 ATTCTGTGATGTTGGTAGGTGCAAGCTTTAGAGCTCTCAGAGAAATGGTGGCAAG 103

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 66 cataattttgagaagccatttggcttgcacatttcttcccataggtgcacacata 125

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 104 GGTCTATTGTGAAAGCCATCTGGTGTGCTGACCTGTAATCATCGGAGAGTTACTAGATG 163

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 126 tcttcttcaaacctttcaggaaagcaaatatagaaattgancatctactaggaggaa 185

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 164 TCTGAAACAGTATCTTTACAGAGAGGAGCAAAATCTTCAGGATTCATCTATTGGGAAAGGA 223

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 186 tncagtnaaatcctncaggttttaaggtttcaaa 221

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 224 GCTTGTGAGAACCTTTCAGTGTCTCGGATTCCTCAAA 259

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4

AV561737

LOCUS

DEFINITION

AV561737

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicaceae; Arabidopsi.

1 (bases 1 to 624)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

AV561737 624 bp mRNA EST 23-JUL-2000

AV561737 Arabidopsis thaliana green siliques Columbia Arabidopsis

thaliana cDNA clone SQ156f07F 3', mRNA sequence.

AV561737

AV561737.1 GI:8733163

EST.

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicaceae; Arabidopsi.

1 (bases 1 to 624)

Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.

COMMENT:

Contact: Smith_TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

```

FEATURES
source      Location/Qualifiers
1..577

```

BASE COUNT	138 a	189 c	143 g	107 t
ORIGIN	lac, hypotalamus, and pituitary."			

	Query Match	16.4%	Score 43.2;	DB 23;	Length 577;
	Best Local Similarity	56.9%;	Pred. No. 0.024;		
	Matches	78;	Mismatches	59;	Indels 0; Gaps 0;
a	45	gaccgcgaagggaatcgcaataatttgagaagccatttggctttgatgcacttc	104		
b	432	gaccacagacggctggaacgcccatcctgtgagaaacccctttggtaggacctgcgagag	491		

RESULT	13
W784451	
OCUS	AW784451
DEFINITION	zbc52g11.g1 Canis cDNAs from mdck cells EST 19-MAY-2000
	zb52g11.5', mRNA sequence.
CCESSION	AW784451
VERSION	AW784451.1 GI:7838827
EYWORDS	EST,
SOURCE	dog

ORGANISM	Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.	
REFERENCE	1 (bases 1 to 385)
AUTHORS	McCombie, W. R., See, L. -H., Baker, J. P., Bahret, A., Dedhia, N. N., de la Bastide, M., Huang, E. N., King, L., Kirchoff, K. A., Miller, B., Nascimento, L. U., O'Shaughnessy, A. L., Preston, R. K., Rodriguez, M. A., Shah, R. S., Shekher, M., Spiegel, L. A., Toth, K., Vil, M. D. and Hannon, G. J.
TITLE	Expressed sequence tags from Canis familiaris (dog)
COMMENT	Unpublished (2000) Contact: W. Richard McCombie Lita Annenberg Hazen Genome Sequencing Center Cold Spring Harbor Laboratory PO Box 100, Cold Spring Harbor, NY 11724, USA Tel: 516 367 8884 Fax: 516 367 8874 Email: mcombie@cshl.org Plate: zb62 row: g column: 11

Query Match	16.2%	Score 42.6;	DB 36;	Length 782;
Best Local Similarity	50.8%;	Pred. No. 0.039;		
Matches	94;	Conservative 0;	Mismatches 91;	Indels 0; Gaps 0;
QY 45	gaccacagaaggatggaatcgcaataattttgagaagccatttgggttttgatgcactttc	104		
Db	504 GACCCAGATAGGCTGGGAACCGCATCATCGTGGAGAAGCCCTTCGGAGGAGGCACCTGCAGAG	563		
QY 105	ttcccataggctgacacaatatctttttcacaacttttcgaaaaagcaaatatatagaat	164		
Db	564 CTCTGACCGGCTGTCCAACACATCTCTCCCTGTTCCGTGAGGACCAGATCTAAGCGAT	623		
QY 165	tgancactactagaaggaatcncagtnaaataacctnccaggttttaagggtttcaaannn	224		
Db	624 CGANCACTACCTGGGCAAGGAGATGGTGAGAACCTTCATGGTGCTGAGATTTGCCAACAGG	683		
QY 225	agntt 229			
Db	684 ATCTTT 688			

Search completed: November 4, 2000, 11:52:01
Job time: 18238 sec

100

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 08:49:11 ; Search time 4075.18 Seconds
(without alignments)
249.718 Million cell updates/sec

Title: US-09-300-482-1

Perfect score: 233

Sequence: 1 gtttttcagtttagtagaat.....ttggacaagggtattgttg 233

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_bal.*

2: gb_ba2.*

3: gb_om.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pl1.*

8: gb_pl2.*

9: gb_pr1.*

10: gb_pr2.*

11: gb_pr3.*

12: em_fun.*

13: em_hum1.*

14: em_hum2.*

15: em_in.*

16: em_om.*

17: em_or.*

18: em_ov.*

19: em_pat.*

20: em_ph.*

21: em_pl.*

22: em_ro.*

23: em_sts.*

24: em_sy.*

25: em_un.*

26: em_vi.*

27: gb_htg1.*

28: gb_htg2.*

29: gb_in1.*

30: gb_in2.*

31: em_bal.*

32: em_ba2.*

33: em_hum3.*

34: em_hum4.*

35: gb_pr4.*

36: gb_htg3.*

37: gb_htg4.*

38: gb_htg5.*

39: gb_htg6.*

40: gb_htg7.*

41: em_htg1.*

42: em_htg2.*

43: em_htg3.*

44: em_hum5.*
45: gb_pl3.*
46: gb_pr5.*
47: gb_htg8.*
48: gb_htg9.*
49: gb_htg10.*
50: gb_htg11.*
51: gb_htg12.*
52: gb_htg13.*
53: gb_htg14.*
54: gb_in3.*
55: gb_htg15.*
56: gb_htg16.*
57: gb_htg17.*
58: em_htg4.*
59: em_htg5.*
60: em_htg6.*
61: em_htg7.*
62: em_hum6.*
63: gb_htg18.*
64: gb_htg19.*
65: gb_ba3.*
66: em_htg8.*
67: em_htg9.*
68: em_htg10.*
69: em_htg11.*
70: em_htg12.*
71: em_htg13.*
72: em_htg14.*
73: em_htg15.*
74: em_htg16.*
75: em_htg17.*
76: em_htg18.*
77: em_htg19.*
78: em_htg20.*
79: em_htg21.*
80: em_htg22.*
81: em_htg23.*
82: gb_pr6.*
83: gb_pr7.*
84: gb_htg20.*
85: gb_htg21.*
86: gb_htg22.*
87: gb_htg23.*
88: gb_ro.*
89: gb_sts1.*
90: gb_sts2.*
91: gb_sy.*
92: gb_un.*
93: gb_vil.*
94: gb_vil2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	131.6	56.5	1904	7	AB029456	AB029456 Triticum
2	131.6	56.5	1924	7	AB029455	AB029455 Triticum
3	131.6	56.5	1956	7	AB029454	AB029454 Triticum
4	106.6	45.8	1785	45	MSU18238	U18238 Medicago sa
5	101.8	43.7	1689	45	STG6PDH	X74421 S.tuberosum
6	100.8	43.3	1957	7	AF012863	AF012863 Petroselin
7	97	41.6	1821	45	NTTCG9	AJ001770 Nicotiana
8	96	41.2	1848	7	AF012862	AF012862 Petroselin
9	95.4	40.9	1852	45	NTTCG6	AJ001769 Nicotiana
10	95.4	40.9	1862	8	ATH010970	AJ010970 Arabidops
11	93.8	40.3	1690	8	ATH010971	AJ010971 Arabidops
12	78	33.5	63604	8	AP000381	AP000381 Arabidops

13	77.8	33.4	1777	7	AF097663	Mesembrya
14	74.6	32.0	14323	72	AB015470	Arabidops
15	39	16.7	37184	82	HSU9D4	
16	39	16.7	180008	48	AC022962	Homo sapi
17	38.4	16.5	171132	50	AC025541	Homo sapi
18	35.8	15.4	147990	51	AC026144	Homo sapi
19	35.8	15.4	181311	64	AL162400	Homo sapi
20	35.4	15.2	208963	84	AL356502	Homo sapi
21	35	15.0	154746	53	AC041038	Homo sapi
22	35	15.0	168832	10	AC007511	Homo sapi
23	34.8	14.9	166391	28	AC009386	Homo sapi
24	34.8	14.9	168006	40	AC019376	Homo sapi
25	34.8	14.9	203894	64	AL162273	
26	34.6	14.8	2945	1	AF269329	Staphyloc
27	34.6	14.8	3825	1	AF269754	Staphyloc
28	34.6	14.8	182737	55	AC068066	Mus muscu
29	33.6	14.4	35739	1	AF015825	Bacillus
30	33.6	14.4	196069	39	AC016722	Homo sapi
31	33.6	14.4	216750	2	BSUB0007	Bacillus su
32	33.2	14.2	150332	9	AC004921	Homo sapi
33	33	14.2	138016	35	CNS01DTP	Human chr
34	33	14.2	175948	40	AC019192	Homo sapi
35	32.8	14.1	128155	46	HSA262A13	Human DNA
36	32.6	14.0	44237	30	CELY45B4B	Caenorhab
37	32.6	14.0	152544	56	AC068750	Homo sapi
38	32.6	14.0	158050	52	AC036237	Homo sapi
39	32.6	14.0	169328	39	AC016281	Homo sapi
40	32.6	14.0	267118	27	AC006889	Caenorhab
41	32.4	13.9	330	90	HSB314WC1	H.sapiens (
42	32.4	13.9	95993	8	ATFE2A19	
43	32.4	13.9	149322	38	AC016039	Homo sapi
44	32.4	13.9	170630	39	AC016905	Homo sapi
45	32.4	13.9	173363	86	AP001989	Homo sapi

ALIGNMENTS

RESULT	1	
AB029456	1904 bp	mrna
LOCUS		04-JUL-2000
DEFINITION	Triticum aestivum G6pdh mRNA for glucose-6-phosphate dehydrogenase, complete cds, clone:Tagp03.	
ACCESSION	AB029456	1 GI:8918505
VERSION	glucose-6-phosphate dehydrogenase.	
KEYWORDS	Triticum aestivum (cultivar:Chinese spring, isolate:root) seedling	
SOURCE	cdNA to mRNA, clone:Tagp03.	
ORGANISM	Triticum aestivum Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum. 1 (sites) Nemoto,Y., Kawakami,N. and Sasakuma,T. Isolation of novel early salt-respnding genes from wheat (Triticum aestivum L.) by differential display Theor. Appl. Genet. 98, 673-678 (1999)	
REFERENCE	2 (sites) Nemoto,Y. Molecular characterization of glucose-6-phosphate dehydrogenase (G6PDH) from wheat (Triticum aestivum L.): gene expression in response to salt stress Unpublished (1999)	
JOURNAL	3 (bases 1 to 1904) Nemoto,Y. Direct Submission	
REFERENCE	Submitted (01-JUL-1999) to the DDBJ/EMBL/GenBank databases. Yasue Nemoto, Yokohama City University, Kihara Institute for Biological Research; Maioaka, Totsuka, Yokohama 244-0813, Japan (E-mail:nemoto@yokohama-cu.ac.jp, Tel:81-45-820-1902, Fax:81-45-820-1901)	
JOURNAL	Location/Qualifiers	
REFERENCE	1. 1904	
AUTHORS	/organism="Triticum aestivum"	
JOURNAL		
FEATURES		
source		

```

/cultivar="Chinese spring"
/isolate="root"
/db_xref="taxon:4565"
/clone="tagpd3"
/dev_stage="seedling"
/note="hexaploid"
153..1682
/gene="g6pdh"
153..1682
/gene="g6pdh"
/codon_start=1
/product="glucose-6-phosphate dehydrogenase"
/protein_id="BAA97664.1"
/db_xref="GI:8918506"
/translation="MAGTDSSEASSRQSFNSLAKDLLEPLKMGCLTIIVVLGASGLAK
KTFPALVHLFQGFLOGEVHIVGATNLSDDLRGRIRAYLKGASEHVFSEFLQL
IKYVSGDTSRVIERKPGDLSAELSLSQELFEEQOLYRIDHYLKGELVONLLVL
RFANRLFLPQNRNDVNIQVIFREDGGVGGYFDQGIIRGLIQHLLQVFLCLA
MEKPVSLKPEHIRDKVKVQLQSVNPKDEEVYQYQGYKEDPTVPDDSNTPFFASIV
LKVHNERWGVFILLKAGKALNSRKAIRVQDPVPGDIFCKKGRNEFVIRLQSP
AMWMLTVKKPKGLEMATQESLDLSVGMRYODVKIPEAYERLILDTIRGQQHFVRRD
ELKAAWQIFTPLLHDDIADGKLKAVSYKPGSRGPKADELSKVGVMQTHGYIWIPTTL
A"
BASE COUNT      543 a   414 c   464 g   483 t
ORIGIN
Query Match      56.5%; Score 131.6; DB 7; Length 1904;
Best Local Similarity 76.3%; Pred. No. 6.3e-29;
Matches 177; Conservative 0; Mismatches 49; Indels 6; Gaps 1;

Qy  2  tttttgcagttagagaattgttagtggtcctctatgattagggtggaagatttgagta 61
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  447 TTCTTGCAATTGATAAATATGATCAGTGGTTCCTATGACAGTGGAGAAGGTTTGAAAA 506
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy  62  ttgaatgaggaactctctgagtagtgagacttcagaaacaataactgactgggaagctaccgc 121
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  507 CTGAACAAGGAATATACAGATTATGAGATGTCA-----AACAACTCAGGAAGTCCTCCGT 560
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy  122 agattattttatttggaatgcctccatcagcttaccatcagtagtcagagatgataaga 181
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  561 AGGCTCTTTTATTGGGCAATTGCCCTCCATCTGTACCCCTTCAGTGTGCAAAATGATCCGA 620
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy  182 tcattatgcagtagtccatcttcacacaccggtgtggacaaggttatgtgt 233
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  621 ACATATTGCATGATGCCAACTCTCCGCGTGGATGGACTAGAGTAATTTGTTG 672
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
AB029455 1924 bp mRNA PLN 04-JUL-2000
LOCUS      Triticum aestivum g6pdh mRNA for glucose-6-phosphate dehydrogenase,
DEFINITION complete cds, clone:tagpd2.
ACCESSION AB029455
VERSION   AB029455.1 GI:8918503
KEYWORDS  glucose-6-phosphate dehydrogenase.
SOURCE    Triticum aestivum (cultivar:Chinese spring, isolate:root) seedling
cDNA to mRNA, clone:tagpd2.
ORGANISM  Triticum aestivum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.
1 (sites)
Nemoto,Y., Kawakami,N. and Sasakuma,T.
Isolation of novel early salt-responding genes from wheat (Triticum
aestivum L.) by differential display
Theor. Appl. Genet. 98, 673-678 (1999)
2 (sites)
Nemoto,Y.
Molecular characterization of glucose-6-phosphate dehydrogenase
(G6PDH) from wheat (Triticum aestivum L.): gene expression in
response to salt stress
Unpublished (1999)
JOURNAL
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE

```

REFERENCE 3 (bases 1 to 1924)
 AUTHORS Nemoto, Y.
 TITLE Direct Submission
 JOURNAL Submitted (01-JUL-1999) to the DDBJ/EMBL/GenBank databases. Yasue Nemoto, Yokohama City University, Kihara Institute for Biological Research; Maiooka, Totsuka, Yokohama 244-0813, Japan (E-mail:nemoto@yokohama-cu.ac.jp, Tel:81-45-820-1902, Fax:81-45-820-1901)

FEATURES
 source
 1..1924
 /organism="Triticum aestivum"
 /cultivar="Chinese spring"
 /isolate="root"
 /db_xref="taxon:4565"
 /clone="tagp2"
 /dev_stage="seedling"
 /note="hexaploid"
 161..1702
 /gene="g6pdh"
 161..1702
 /gene="g6pdh"
 /codon_start=1
 /product="glucose-6-phosphate dehydrogenase"
 /protein_id="BAA97863.1"
 /db_xref="GI:8918504"
 /translation="MAGTDSASSRSQSFNSLAKDLPLEQGLTIIVLGASGLAK KKTFFALYHLFEQGLQSGEVHIVGYARTNLSDGLGRIRAYLKGASEHVSFQLQ IKVSGSYDSGEGEFKLNKEISDYEMSNSSGSRRLFYALPPSVYPSVCKMIRTYCM SPTSRAGWTRVIVEKPFGRDLDSAEELSSOLGELFEDOLYRIDHYLKGELVONLLVL RFANRLFLPNRNDVNIQIVFREDFGTGRGQYFDQYGIIRDIQNHLLQVCLVA MEKPSLKEPHIRDEKVLQSNVPIKDEEVQLQGVQKDDPTVPDSDNTPTFASIV LRVHNERWGPVFLKAGKALNSRAEIRVQFKDVPGLDKCKKQGRNEFVIRLOPSE AMYMKLTFVQVKKPGLEMAEQSELDLSYGMRYQDVKIPEAYERLLIDTIRGDOOHF VRDELKAAWQITPPLLHDIDAGKLKAVSYKPGSRKPEADELSEKVGVMQTHGYIWI PPILA"

BASE COUNT 545 a 417 c 475 g 487 t
 ORIGIN

Query Match 56.5%; Score 131.6; DB 7; Length 1924;
 Best Local Similarity 76.3%; Pred. No. 6.3e-29;
 Matches 177; Conservative 0; Mismatches 49; Indels 6; Gaps 1;

QY 2 ttttgcagttagagaataatttagtggctctcatgtagggtggaagattgagttta 61
 Db 455 TTCTTGCAATTGATAAATATGTCAGTGGTTCCTATGACAGTGGAGAGTTTGA 514
 QY 62 ttgaatgagcaatctctgagtagtagacttcagaaacaatgactcgggaagctaccgc 121
 Db 515 CTGAACAAGGAATATCAGATTATGAGATGTCA-----AACAACTCAGGAAGCTCCCGT 568
 QY 122 agattatttttggcattgctccatcagttcaccatcagtagtcgagatgataaga 181
 Db 569 AGGCTCTTTATTGGCATTGCTCTCATCTGCTACCTTCAGTGTGCAAAATGATCCGA 628
 QY 182 tcatattgcatgagtcacattcttcacacacccggttggaaggttattgttg 233
 Db 629 ACATATTGCATGAGTCCCACTTCTCGCGTGGATGACTAGAGTAATTGTTG 680

RESULT 3
 LOCUS AB029454
 DEFINITION Triticum aestivum g6pdh mRNA for glucose-6-phosphate dehydrogenase, complete cds, clone:Tagpdl. 04-JUL-2000
 ACCESSION AB029454
 VERSION AB029454.1
 KEYWORDS glucose-6-phosphate dehydrogenase.
 SOURCE Triticum aestivum (cultivar:Chinese Spring, isolate:root) seedling
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Triticum.

REFERENCE 1 (sites)
 AUTHORS Nemoto, Y., Kawakami, N. and Sasakuma, T.
 TITLE Isolation of novel early salt-responding genes from wheat (Triticum aestivum L.) by differential display

JOURNAL Theor. Appl. Genet. 98, 673-678 (1999)
 REFERENCE 2 (sites)
 AUTHORS Nemoto, Y.
 TITLE Molecular characterization of glucose-6-phosphate dehydrogenase (G6PDH) from wheat (Triticum aestivum L.): gene expression in response to salt stress

Unpublished (1999)
 3 (bases 1 to 1956)
 Nemoto, Y.
 Direct Submission
 Submitted (01-JUL-1999) to the DDBJ/EMBL/GenBank databases. Yasue Nemoto, Yokohama City University, Kihara Institute for Biological Research; Maiooka, Totsuka, Yokohama 244-0813, Japan (E-mail:nemoto@yokohama-cu.ac.jp, Tel:81-45-820-1902, Fax:81-45-820-1901)

FEATURES
 source
 1..1956
 /organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /isolate="root"
 /db_xref="taxon:4565"
 /clone="tagpdl"
 /dev_stage="seedling"
 /note="hexaploid"
 162..1691
 /gene="g6pdh"
 162..1691
 /gene="g6pdh"
 /codon_start=1
 /product="glucose-6-phosphate dehydrogenase"
 /protein_id="BAA97862.1"
 /db_xref="GI:8918502"
 /translation="MAGTDSASSRSQSFNSLAKDLPLEQGLTIIVLGASGLAK KKTFFALYHLFEQGLQSGEVHIVGYARTNLSDGLGRIRAYLKGASEHVSFQLQ IKVSGSYDSGEGEFKLNKEISDYEMSNSSGSRRLFYALPPSVYPSVCKMIRTYCM SPTSRAGWTRVIVEKPFGRDLDSAEELSSOLGELFEDOLYRIDHYLKGELVONLLVL RFANRLFLPNRNDVNIQIVFREDFGTGRGQYFDQYGIIRDIQNHLLQVCLVA MEKPSLKEPHIRDEKVLQSNVPIKDEEVQLQGVQKDDPTVPDSDNTPTFASIV LRVHNERWGPVFLKAGKALNSRAEIRVQFKDVPGLDKCKKQGRNEFVIRLOPSE AMYMKLTFVQVKKPGLEMAEQSELDLSYGMRYQDVKIPEAYERLLIDTIRGDOOHF ELKAAWQITPPLLHDIDAGKLKAVSYKPGSRKPEADELSEKVGVMQTHGYIWIPTLA"

BASE COUNT 592 a 413 c 465 g 486 t
 ORIGIN

Query Match 56.5%; Score 131.6; DB 7; Length 1956;
 Best Local Similarity 76.3%; Pred. No. 6.3e-29;
 Matches 177; Conservative 0; Mismatches 49; Indels 6; Gaps 1;

QY 2 ttttgcagttagagaataatttagtggctctcatgtagggtggaagattgagttta 61
 Db 456 TTCTTGCAATTGATAAATATGTCAGTGGTTCCTATGACAGTGGAGAGTTTGA 515
 QY 62 ttgaatgagcaatctctgagtagtagacttcagaaacaatgactcgggaagctaccgc 121
 Db 516 CTGAACAAGGAATATCAGATTATGAGATGTCA-----AACAACTCAGGAAGCTCCCGT 569
 QY 122 agattatttttggcattgctccatcagttcaccatcagtagtcgagatgataaga 181
 Db 570 AGGCTCTTTATTGGCATTGCTCTCATCTGCTACCTTCAGTGTGCAAAATGATCCGA 629
 QY 182 tcatattgcatgagtcacattcttcacacacccggttggaaggttattgttg 233
 Db 630 ACATATTGCATGAGTCCCACTTCTCGCGTGGATGACTAGAGTAATTGTTG 681

RESULT 4
 MSU18238

```

LOCUS      MSU18238      1785 bp      mRNA      PLN      30-JAN-1997
DEFINITION Medicago sativa glucose-6-phosphate dehydrogenase mRNA, complete
            cds.
ACCESSION  U18238
VERSION    U18238.1   GI:603218
KEYWORDS   .
SOURCE     alfalfa.
ORGANISM   Medicago sativa subsp. sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE  1 (bases 1 to 1785)
AUTHORS   Fahrendorf,T., Ni,W., Shorosh,B.S. and Dixon,R.A.
TITLE     Stress responses in alfalfa (Medicago sativa L.) XIX.
            Transcriptional activation of oxidative pentose phosphate pathway
            genes at the onset of the isoflavonoid phytoalexin response
            Plant Mol. Biol. 28 (5), 885-900 (1995)
JOURNAL   95367649
MEDLINE   2 (bases 1 to 1785)
REFERENCE 2
AUTHORS   Fahrendorf,T.
TITLE     Direct Submission
            Submitted (06-DEC-1994) Theo Fahrendorf, Samuel Roberts Noble
            Foundation, Plant Biology Division, 2510 Sam Noble Parkway,
            Ardmore, OK 73402, USA
JOURNAL
FEATURES   Location/Qualifiers
            1..1785
            /organism="Medicago sativa subsp. sativa"
            /cultivar="Apollo"
            /sub_species="sativa"
            /db_xref="taxon:56147"
            /clone="G10"
            /tissue_type="cell suspension culture"
            38..1585
            /codon_start=1
            /product="glucose-6-phosphate dehydrogenase"
            /protein_id="AAB41552.1"
            /db_xref="GI:603219"
            /translation="MGTNVHVERRDSIGTESPVAREVLETGTSIVVLGASGDGLAKK
            KTFPALFLYKQELLPPDEHIFGYARKSIDDELNRKLSRYLVPEKGASPKOLDVVS
            KFLQLVKYQSEDPDEGFRLLDKSEHEVLYKNSKEGSRRLFLYALPSPVPSVCK
            MIKCCMKNSDLGWRTRVVEKPGROLESARELSSTOIGELFEPOLYRIDHLYGREL
            VONMLVFAFRFLPLMNNHIDNVQIVFREDEGTGREGGYDQYGIIRDILPNHLL
            QVCLIAKEKPVSLKPEHIDEKVKLVLESVLPFRIDDEVILQYEGYDIDPTVPDDSN
            TPTATTILIRINERWEGVPFVFKAGKALNSKAEIRVQFKDVPDGFIRSKKQGRNEFV
            IRLOPSEAIYMKLVKQPLEMSAVQSELDLSYQRYQGIIPEAYERLILDTIRGDDQ
            OHFVRDELKASQWIFTPLLHKIDRGELKPVYPNPGSRPAEADDELLEKAGYVQTPGY
            IWIPPTL"
            1785
            /note="13 A nucleotides"
            551 a 306 c 396 g 532 t
            polyA_site
BASE COUNT 551 a 306 c 396 g 532 t
ORIGIN
Query Match 45.8%; Score 106.6; DB 45; Length 1785;
Best Local Similarity 66.1%; Pred. No. 1.8e-21;
Matches 154; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 1 gtttttcagttagtagaataatttagtgctcctctatgatagggtggaaggatttgatt 60
Dbb 346 GTTTTACAAATGTTAAATGTAAGTGCCCTTAAGATTCTGAAGATGGATTTCGCTT 405
QY 61 attgaatgagcgaatctctgagtagatgagacttcagaaacaatgactcgggaagctaccg 120
Dbb 406 GTTGGATAAGAGATTTCAGAGCATGAATATTGAAAAATAGTAAAGAGGGTTTCATCTCG 465
QY 121 cagattattttattggcattgctccatcattcaccatcagtagtgcgaatgataag 180
Dbb 466 GAGGCTTTTCTATCTGCATCTCCTCCTCAGTGTATCATCGTTTCGAAGATGATCAA 525
QY 181 atcatattgagtagccatctccacacacccggtttggcaagsggtttattgttg 233
Dbb 526 AACTTGTTGCATGAATAAATCTCATCTTGTGATGCACACGCGTTGTTGTTG 578

```

```

RESULT      5
STG6PDH
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES
source
Location/Qualifiers
1..1689
/organism="Solanum tuberosum"
/cultivar="Desiree"
/db_xref="taxon:4113"
/haplotype="diploid"
/dev_stage="vegetative"
/tissue_type="green leaves"
/cell_type="mesophyll"
/clone_lib="lambda ZAPII"
/clone="K4"
46..1581
/gene="g6pdh"
46..1581
/EC_number="1.1.1.49"
/note="homotetramer"
/codon_start=1
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="CAA52442.1"
/db_xref="GI:471345"
/db_xref="SWISS-PROT:P37830"
/translocation="MAASWCIKRGSRINDSPNDNIPETGCLSTIVLGSGLAKK
KTFPALFLYKQFLOSNEVHIFGYARKSIDDELNRKLSRYLVPEKGASPKOLDVVS
LIRYVSGYSDSAGEFTSLDKAISSEHFSKNSGSRRLFLYALPSPVPSVCMIKS
YCNKSDLGQWTRTVYKPKGLASSELSSQIGELFDEPOIYRIDHLYGRELVLN
LVLRANRFFPLNWRNDINQIVFREDFGTGREGGYDEYGIIRDILHLLQVLC
LVAMEKPVSKPEHIDEKVKLVQSMPLTEDEVILQYEGYDIDPTVPDDSN
TWVLRINERWESVPIFMKAGKALNSKAEIRVQFKDVPDGFIRCKQGRNEFVIRLO
PSAMYNKLVTKKPGLEMSVQSELDLSYQRYQGVVPEAYERLILDTIRGDDQHFV
RDELKAAWEIFTPLLRINDNGEKPIPYKPGSRPAEADDELLEQNGAGYVQTHYIWP
PTL"
BASE COUNT 488 a 317 c 393 g 491 t
ORIGIN
Query Match 43.7%; Score 101.8; DB 45; Length 1689;
Best Local Similarity 64.8%; Pred. No. 4.9e-20;
Matches 151; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
QY 1 gtttttcagttagtagaataatttagtgctcctctatgatagggtggaaggatttgatt 60
Dbb 342 GTTTCGCAACTGATAAATACGAGTGTCTTATGATTCTCTGAGGATTTAGTCTC 401
QY 61 attgaatgagcgaatctctgtagtagatgagacttcagaaacaatgactcgggaagctaccg 120

```

402 ACTGGATAGGCAATATCTGACGCGAATCTCAAAAATAGCACAGGATCATCCAG 461
 121 cagatattttatggcattgcctccatcagtcctacccatcagtcgcgaataataag 180
 462 AAGACTCTCTCTATTGTGCTTCTCCATCAGATATATCCCTCTGTGTGGCAGATGATAA 521
 181 atcatattgcagtcagtcacattccacacccggttggacaaagggttattggtg 233
 522 AAGTTATTGTATGAACAAATCTGATCTTGTTGTTGGACACGGACATGTTGTG 574

RESULT	6
AF012863	
LOCUS	1957 bp mRNA PLN 03-SEP-1997
DEFINITION	Petroselinum crispum cytosolic glucose-6-phosphate dehydrogenase 2 (cg6PDH2) mRNA, complete cds.
ACCESSION	AF012863
VERSION	AF012863.1 GI:2352922
KEYWORDS	.
SOURCE	parsley.
ORGANISM	Petroselinum crispum Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum. 1 (bases 1 to 1957)
REFERENCE	Batz,O., Logemann,E., Reinold,S. and Hahlbrock,K. Extensive reprogramming of cellular metabolism by fungal elicitor or infection in parsley suggests a new perception of 'defense-related' genes Unpublished
JOURNAL	2 (bases 1 to 1957) Batz,O., Logemann,E. and Hahlbrock,K. Direct Submission
REFERENCE	Submitted (08-JUL-1997) Biochemistry, MPI f. Zuechtungsforschung, Carl-von-Linne-Weg 10, Cologne, NRW 50829, Germany Location/Qualifiers 1..1957
AUTHORS	/organism="Petroselinum crispum" /db_xref="taxon:4043"
TITLE	/cell_type="cultured cells"
JOURNAL	1..1957
FEATURES	74..1678 /gene="cg6PDH2" /gene="cg6PDH2" /codon_start=1 /product="cytosolic glucose-6-phosphate dehydrogenase 2" /protein_id="AAB69319.1" /db_xref="GI:2352923"
source	/translation="MKSAVCDDIVHKMSDKWCLERAPWVEDSGADPTKDVTOTPPPVV EDGCLSVNVLGASGLAKKVFPAFLNLRYRGFQGSHEVIIFGYARTKIISDDLDRRI RGTLPNNKTAGHTEDYSKFLLIKYVCGSYDAEGFCQLDREISEHSIRNSIEGSI SRRLFYALPPSPVPCKMLTKTYCMKNKSLDGWTIRIVEKPFRDLSEAEHLNQIG ELFEPEQIYRDHYLKGELVNOLILVRPANFFPLPSWNRNDSISSVOIVREFDFGTGEI GYTFOYGIRDIIONHULLVLCIVAMEKPISLKPEHRIDEKVKVLQSVAFLNDEEV LGQYDGYLDPTFPVAGSTPTTFATVMILRHNERGWPFVLKACKALESKRAEIQTQ KQGVGDIFCKQKQNEFVIRLQSEAMYMKLTVKKPGLEMWSTAQSELDLSVGORYOD VTIPAYERLLIDITIRGQQHFVRDELKAWEIFTLLHRIDKSELKSVSYPKPGSRG PBEADELLKKVGYVTHCYTWTPTL"
gene	571 a 361 c 451 g 574 t
CDS	
BASE COUNT	
ORIGIN	

	Query Match	43.3%	Score 100.8;	DB 7;	Length 1957;
	Best Local Similarity	64.7%	Prod. No. 9.7e-20;		
	Matches 150;	Conservative	0;	Mismatches 82;	Indels 0; Gaps 0;
QY	2	tttttcgattgtagaataatgtagctcctatgtagggtggaaggatttgagtta	61		
Db	440	TTTCTGCAACTGATCAAAATACGTCCTGTGGGTCTTATGACGCAGCGAGGGGTTTTCAGTGT	499		
QY	62	ttagaatgagggaactctctcgatgtagactgcagaaaaaatgactcgggaagctaccgc	121		

500	TTGGATAAGGAATATCTGAGCATGAAATATCAAGAAACAGTATTGAAGGATCGTCCCGG	559
Qy		
122	agattattttatggcattgctcccatcagctaccocatcagtatcgcgagatgataaga	181
Db		
560	AGCATATTTTACCTTGGCACTTCCCTTCCTCAGTATATCCTCTGCTGC AAAATGATTAA G	619
Qy		
182	tcattatgcatgagtcctatcttcacacacccggttggacaaggggtattgttg	233
Db		
620	ACTTACTGCATGAATPAATATCGATCTTGGTGGATGGACTCGTATTTGTTGTG	671
Qy		

RESULT	7	
NTTCG9		
LOCUS	1821 bp	mRNA
DEFINITION	17-AUG-1999	PLN
		Nicotiana tabacum mRNA for cytosolic glucose-6-phosphate dehydrogenase TCG9.
ACCESSION	AJ001770	
VERSION		
KEYWORDS	AJ001770.1	GI:3021509
SOURCE		glucose-6-phosphate dehydrogenase.
ORGANISM		common tobacco.
		Nicotiana tabacum
		Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridaceae; euasterids I; Solanales; Solanaceae; Nicotiana
REFERENCE		1 (bases 1 to 1821)
AUTHORS		Wendt,U.K., Hauschild,R., Lange,C., Pietersma,M., Wenderoth,I. and von Schaewen,A.
TITLE		Evidence for functional convergence of redox regulation in G6PDH isoforms of cyanobacteria and higher plants
JOURNAL		Plant Mol. Biol. 40 (3), 487-494 (1999)
MEDLINE		99364543
REFERENCE		2 (bases 1 to 1821)
AUTHORS		von Schaewen,A.
TITLE		Direct Submission
JOURNAL		Submitted (18-SEP-1997) von Schaewen A., Plant Physiology, University Osnabrueck, Barbarastr. 11, D-49069 Osnabrueck, GERMANY
FEATURES		Location/Qualifiers
source		1..1821

```

CDS
/tissue_type="young sink leaves"
88..1623
/note="TCG9"
/codon_start=1
/product="glucose-6-phosphate dehydrogenase"
/protein_id="CAA04993.1"
/db_xref="GI:3021510"
/db_xref="SPTREMBL:O65855"
/translation="MAASNCIEKRGSIRLDSFRDNINPTEGCLSIIVLASGDLAKK
KTFPLNFYLRQGFQSNVHIFGYARTKIISDDLLRGYRLSQSGNEEYSEFIQ
LIKYVSGDYSGEGESLLDKATARIEAKNTEGSSRRFLFYALPQSVPCVMKIL
YCNKNSDLGGWTRIVVEKPEFGDLASQLSQIGELFDEFOIYKIDHYLGKELQVNL
LYLRFANREFLPLNRNDINDNTIQIVREDFTEGRCGYDEYGIIRDITIQNLQLVL
LYAMEKPVSKQEHILNDEKVAQLQSMPLKIDEEVVLGGYEGYKDDTPVPNSNPTFA
TWLIRNMTKRWGSGVFKMKAGALNSAKRIYQFQKDPGIFRCKQGNREYFIRLQ
PEAMTYKLTIVKKGPLEMSTVQSELDLSYRQYQGVIPAYERLIIDTIRGQOQHFV
RDELKAAWEIETPLLLRHIDDEGVKPIPKYKPSRGPAEADELLQNVGVYVQTHGVICIP
PTL"
515 a 340 c 434 g 532 t
BASE COUNT
ORIGIN

```

```

Query Match      41.6%  Score 97;  DB 45;  Length 1821;
Best Local Similarity 53.5%  Pred. No. 1.3e-18;
Matches 148;  Conservative 0;  Mismatches 85;  Indels 0;  Gaps 0;

Oxy 1 gtttttcgagtttagaataatgttagtgctgctctatgaggtgggaaggatttgagt 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
db 384 GTTCTCTCAGCTGATTAAATACGCTCAGTGCCTCTTATGATCTGGGGAGGCGCTTAGTT 443

```


Qy	61	atgaatgaggcaatctctgagtatgagacttcagaaaaacaatgactcggaagctaccg	120
Db	455	ACTGACAAAGCGGATAGCTGAGCACGAATTCGCAAAAATACACGGAAGGATCATCCAG	514
Qy	121	cagattatttattgacattgctccatcgaatcagtcattcccatcagtcgagatgataag	180
Db	515	AAGACTCTTACATTGCTCTTCCCTCCATCAGTATATCCCTCTGTTTCAGGATGATAAA	574
Qy	181	atcatattgatcagtcattcattccacacaccggttggacaagggttatgttg	233
Db	575	AAACTATTGATGAACAAATCTGATCTTGGTGGTGGATCGCATTTGTTTG	627
RESULT	10		
LOCUS	ATH010970		
DEFINITION	Arabidopsis thaliana mRNA for cytosolic glucose-6-phosphate	17-AUG-1999	
ACCESSION	AJ010970		
VERSION	AJ010970.1	GI:5732194	
KEYWORDS	acg9 gene; glucose-6-phosphate 1-dehydrogenase.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1 (bases 1 to 1862) Wendt, U.K., Hauschild, R., Lange, C., Pietersma, M., Wenderoth, I. and von Schaeuwen, A.		
TITLE	Evidence for functional convergence of redox regulation in G6PDH isoforms of cyanobacteria and higher plants		
JOURNAL	Plant Mol. Biol. 40 (3), 487-494 (1999)		
MEDLINE	99364543		
REFERENCE	2 (bases 1 to 1862) von Schaeuwen, A.		
AUTHORS	Direct Submission		
TITLE	Submitted (04-SEP-1998) von Schaeuwen A., Plant Physiology, University Osnabrueck, Barbarastrasse 11, D-49069 Osnabrueck, GERMANY		
JOURNAL			

FEATURES	GENOTYPE	Location/Qualifiers	FEATURES	JOURNAL
source	1. .1862	/organism="Arabidopsis thaliana"	source	
gene	43..1503	/db_xref="taxon:3702"		

gene	43. .1593	
CDS	/gene="acg9"	
	43. .1593	
	/gene="acg9"	
gene	/EC_number="1.1.1.49"	
CDS	/codon_start=1	

```

/cocon_start=1
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="CAB52674.1"
/db_xref="GI:5732195"

```

translation="MGSQGWHEKRSKTIANDSFVEYKYNPIETETSGLSIIVLGASGDILA
KKTPPALFNLHQIQLNDVEVHIFGIAKSLITDEELDKIRGLYLVPEKNAKSKTEAL
SKFLIKYVSGPYDSEEGFKRLDNLALHEHESKTKATSGRRSLYVIALPSPVVPVS
KMLKAMGTANSGDGLGWTRVLPVPEKFDKLSAEUSSQIGALFEPQYIYDIHYLGKE
LVONNLVLPANFLPFLPLWRNDRINANVLVFEDETEGREGYDFEYIIRLQIHLH
LOVLCIVAMKEKISLAPHEIRIOEKVQLQSVIPIKDEEVLQGYEGYRDDTPVNSD
TPPTATLILKINNEWGVFFILKAKGAMSKKADIIRQKDFVDFIDQKQNGRNEF
VQIOPSEAMKMLTKVQPLEMOTVQSELDLSYKQYQDIPSEAYEALILUOTIRGD
CHQFVRDEUKAAMEFTPLLRHDKGVSVYPIKQSGRPAEQALLKKAGYMYOHTG

BASE COUNT	564 a	330 c	420 g	548 t
ORIGIN				

Query Match 40.9%; Score 95.4; DB 8; Length 1862;
Best Local Similarity 63.1%; Pred. NO. 4e-18;
Matches 147; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1 gttttgcagtttagtagaataatgttagtggtcctcctatgatagggtggaaggatttgagtt 60

Db	354	GTTTCTAAAGCTGATTAAAGTATGTGAGTGGACCTTATGATCTGAGGAAGGTTTAAAG	413
Qy	61	attgaatgaggaactctcgtgatgagactcagaaaaaatgaactcgtcggaagctaacg	120
Db	414	ATTAGACAAGCGCATTTTGGAGCACGAGATATCTAAAAGACTGCTGAAGGATCTTCGAG	473
Qy	121	cagattatttatttggcattgctccatcagctaccocatcagttatcgagatgataag	180
Db	474	GAGATTGTTTATCTTTGCACTTCTCCGCTCTATACCTCTCTGAAGCAAGATGATCAA	533
Qy	181	atcattatgatgatgcattcttcacacacccggttggacaagggtattgttg	233
Db	534	GGCATGGTGCATTAACAAATCTGATCTTGGTGGATGGAGCTAGGATTTGTTTG	586

RESULT	11
ATH010971	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eumliophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
1 (bases 1 to 1690)
Wendt, U.K., Hauschild, R., Lange, C., Pietersma, M., Wenderoth, I. and von Schaewen, A.
Evidence for functional convergence of redox regulation in G6PDH isoforms of cyanobacteria and higher plants
Plant Mol. Biol. 40 (3), 487-494 (1996)
93364543

REFERENCE
2 (bases 1 to 1690)
von Schaewen, A.
Direct Submission
Submitted (04-SEP-1998) von Schaewen A., Plant Physiology,
University Osnabrueck, Barbarastrasse 11, D-49069 Osnabrueck,
GERMANY

FEATURES
SOURCE

source	gene	CDS
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```

07.1014
/gene="acg12"
/EC_number="1.1.1.49"
/codon_start=1
/product="glucose-6-p
/protein_id="CAB52675
/db_xref="GI:5732107"

```

/translation="MSGQWHEKSTRFNSFVREYIVPCTGCLSIIVLGASGDLA
KKTFFALFNROQFLPDEVEHYFATKISDEELNDRFVLDVKNQAEALS
KFLQLKVVSGYPDAEGROUDKAI SEHEITSKCTEGSSRFLYALPSPVPSVCK
MIKCTCMNKSDLGGWTRIVVEKPFCKDLSEAEQLSSQIGELFDESQIHLVGLKEL
VOMNLYLRANRFPFLPNRDNENQVIVREDFETEGRGGYFDEYIITDITONHLL
QVCLVMAKEPISLAKPEHRDSEKVLQVFSRPIVSDVEVLGQYEGYRDDTVPFNSD
PFTATILIRHNEQYGFPLKANGLSNRKAEIRIQFDKVPGDIFRCQKNGRNFV
IRLOPEAMKMLTVKQGPOLMVSLELDLSYGORYQVGAIPYAEVRLTLDITKGDQ
QHVFVRDELKVAWEIFTPLLRIDKGEVKSIPYKPGSRGPKQADLLEKASYLQTHGY
IWIPTPL"

BASE COUNT	506 a	318 c	390 g	476 t

Query Match 40.3%; Score 93.8; DB 8; Length 1690;
Best Local Similarity 62.7%; Pred. NO. 1.2e-17;
Matches 146; Conservative 0; Mismatches 87; Indels 0

0:

Query Match	33.4%	Score 77.8;	DB 7;	Length 1777;	
Best Local Similarity	65.2%	Pred. No. 7e-13;			
Matches 131;	Conservative	0;	Mismatches 67;	Indels 3;	Gaps 1;
QY	2	tttttcgagttagtagaatattgtagtgcctctatgataggggtgaagattgagttga 61			
DB	357	TTTCTGCATCTGATTAAATATGTTAGTGGCTTTATGACACTGAGGACGGTTTCACTTG 416			
QY	62	ttagatagggaactctctgagtagtagaacttcagaataaactcgggaagctaccgc 121			
DB	417	TTAGATGATGAATATCAATCAAGTATGAG---TGTCAAAGAACAACGACGGAACATTTCTAAG 473			
QY	122	agattatttttggcattgcctccatcagctctaccatcagtagtcgagatgataaga 181			
DB	474	AGACTCTTTTATCTTGCTCTTCTCCATCAGTATATCCATCAGTCTGCAGATGATAAAG 533			
QY	182	tcatttgcgatgagtcacatct 202			
DB	534	CTTTGCTGCTATGAATAAATCT 554			
RESULT	14				
AB015470					
LOCUS	AB015470	14323 bp	DNA	PLN	09-AUG-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K1B16.				
ACCESSION	AB015470				
VERSION	AB015470.1	GI:3241918			
KEYWORDS	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC				
SOURCE	Clone:K1B16.				
ORGANISM	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (sites)				
AUTHORS	Nakamura,Y., Sato,S., Asamizu,E., Kaneko,T., Kotani,H., Miyajima,N. and Tabata,S.				
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. VII. Sequence features of the regions of 1,013,767 bp covered by sixteen physically assigned P1 and TAC clones				
JOURNAL	DNA Res. 5 (5), 297-308 (1998)				
MEDLINE	99087489				
REFERENCE	2 (bases 1 to 14323)				
AUTHORS	Nakamura,Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUN-1998) to the DDBJ/EMBL/GenBank databases. Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)				
COMMENT	Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K1B16 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge-National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplinePredictor (Volker Brendel, Stanford University, http://genome1.zool.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MNF13 and the 3' clone is MHK7. Location/Qualifiers 1. .14323				
FEATURES	SOURCE				
1	14323				


```
repeat_region /note="L1 element fragment"
35350..35630
repeat_region /note="L1 element fragment"
36315..36395
repeat_region /note="L1 element fragment"
36510..36657
repeat_region /note="L1 element fragment"
36700..36819
repeat_region /note="MER42A element fragment"
36705..36785
repeat_region /note="MER42C element fragment"
36841..37025
repeat_region /note="L1 element fragment"
37026..37184
repeat_region /partial
/note="Alu repeat: matches 1..174 of consensus"
BASE COUNT 12421 a 7079 c 6961 g 10721 t 2 others
ORIGIN

Query Match 16.7%; Score 39; DB 82; Length 37184;
Best Local Similarity 50.3%; Pred. No. 0.28;
Matches 96; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1 gtttttcagtttagaataatgttagtggtcctctatgataggggtggaaggatttgatt 60
Db 8340 GCTTCTGCATTTGTAGATCATATATATCAACTATATTAATAACAGAGTGAGAGTCTCAA 8281

QY 61 attgaatgaggcaatctctgagtagatgagacttcagaaaaacaatgactcgggaagctaccg 120
Db 8280 ATAGCTAGCTGACATCTCAGTGGACTAGCTATCAACATTTGCTAGCTCTGTAGCATGGA 8221

QY 121 cagattattttatttggttcctccatcagtcctaccatcagtagatgcgagatgataag 180
Db 8220 CAGATTATTTCAGTTATCTGTCTCTCTTTATTTCCTGTAAACATGGAGAGATAATATTAG 8161

QY 181 atcatatigca 191
Db 8160 TGCCTATCTCA 8150
```

Search completed: November 4, 2000, 13:30:03
Job time: 16852 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 09:11:13 ; Search time 320.8 Seconds
(without alignments)
272.847 Million cell updates/sec

Title: US-09-300-482-1

Perfect score: 233

Sequence: 1 gttttgcagtttagtaaat.....ttggacaagggttattgttg 233

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

1: /SID56/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID56/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID56/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID56/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID56/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID56/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID56/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID56/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID56/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID56/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID56/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID56/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID56/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID56/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID56/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID56/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID56/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31.6	13.6	278	21	A01024
2	31.2	13.4	4418	18	W47465
3	29.8	12.8	5280	18	W74855
C 4	29.8	12.8	5631	20	X85506
C 5	29.8	12.8	7655	16	O86896
C 6	29.2	12.5	5342	21	D00374
C 7	28.8	12.4	5354	20	X13263
C 8	28.6	12.3	18613	18	V74423
C 9	28.4	12.2	7313	19	V29268
C 10	28.4	12.2	3508	16	O86693
C 11	28.4	12.2	5333	18	W74481
C 12	28.4	12.2	17710	19	V31256

C 13	28.2	12.1	11887	19	V52279
C 14	28.2	12.1	12565	19	V62392
C 15	28.2	12.1	12565	20	X75925
C 16	28	12.0	337	17	Q59415
C 17	28	12.0	3417	17	T39622
C 18	28	12.0	3417	20	X82491
C 19	27.8	11.9	3050	20	X29768
C 20	27.6	11.8	32768	19	V52204
C 21	27.4	11.8	527	19	V31323
C 22	27.4	11.8	1172	19	V53308
C 23	27.4	11.8	1513	21	A26678
C 24	27.4	11.8	2442	20	Z23609
C 25	27.4	11.8	5892	18	T79680
C 26	27.4	11.8	7240	18	T79680
C 27	27.4	11.8	8703	19	V52310
C 28	27.4	11.8	10485	20	X30256
C 29	27.4	11.8	10485	20	X30257
C 30	27.4	11.8	10485	20	X30258
C 31	27.4	11.8	10485	20	X30259
C 32	27.4	11.8	10485	20	X30255
C 33	27.4	11.8	11283	18	T84841
C 34	27.4	11.8	11283	21	T87986
C 35	27.4	11.8	11385	18	T69707
C 36	27.2	11.7	439	21	A32071
C 37	27.2	11.7	860	18	X30792
C 38	27.2	11.7	1153	19	T98660
C 39	27.2	11.7	3764	21	Z44192
C 40	27	11.6	10240	19	V39007
C 41	26.8	11.5	690	21	A01923
C 42	26.8	11.5	1673	20	X77320
C 43	26.6	11.4	297	20	Z12750
C 44	26.6	11.4	738	20	Z15624
C 45	26.6	11.4	738	20	Z15625

ALIGNMENTS

RESULT 1

A01024
ID A01024 standard; cDNA; 278 BP.

XX A01024;

XX A01024;

DT 19-MAY-2000 (first entry)

XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:1015.

XX Human; colon cancer; tumour; diagnosis; gene expression product;

XX probe; detection; cancerous state; metastasis; identification;

XX breast cancer; oestrogen receptor-positive breast cancer; therapy;

XX oestrogen receptor-negative breast cancer; lung cancer; ss.

XX Homo sapiens.

XX WO9958675-A2.

XX 18-NOV-1999.

XX 13-MAY-1999; 99WO-US10602.

XX 14-MAY-1998; 98US-0085426.

XX 15-MAY-1998; 98US-0085537.

XX 15-MAY-1998; 98US-0085696.

XX 21-OCT-1998; 98US-0105234.

XX 27-OCT-1998; 98US-0105877.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

XX Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

XX Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

Streptococcus pneu
Human interleukin-
Human interleukin
Human brain Expre
Human DNA ligase I
Human DNA ligase I
L.lactis HsdR gene
Streptococcus pneu
E. coli J96 pathog
DNA encoding a Sta
Candida albicans p
Human labyrinthin
BRCA2 cancer susce
Partial BRCA2 can
Streptococcus pneu
Human BRCA2 (om12)
Human BRCA2 (om13)
Human BRCA2 (om14)
Human BRCA2 (om15)
Human BRCA2 (om11)
Human breast and o
BRCA2 gene sequenc
Human breast cance
Plant microsatelli
Streptococcus pneu
DNA encoding a S.
Murine cerebral ne
Porcine INF-alpha
Human colon cancer
Human secreted pro
Human gene express
Human gene express
Human gene express

FT misc_feature 13501..13560
 FT /*tag= h
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 15301..15360
 FT /*tag= i
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 FT misc_feature 17101..17160
 FT /*tag= j
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT
 PN EP786519-A2.
 XX
 XX 30-JUL-1997.
 XX
 XX 07-JAN-1997; 97EP-0100117.
 XX
 XX 05-JAN-1996; 96US-0009861.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 XX
 XX WPI; 1997-374922/35.
 XX
 XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 XX
 XX Claim 1; Page 641-651; 3271pp; English.
 XX
 CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 XX
 XX Sequence 18613 BP; 5990 A; 2753 C; 3477 G; 5782 T; 611 other;
 SQ
 Query Match 12.4%; Score 28.8; DB 18; Length 18613;
 Best Local Similarity 48.2%; Pred. No. 10;
 Matches 81; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 58 gttattgaatgaggaactctctgagtatgagactcagaacaaatgactcgggaagcta 117
 DB 18329 GTTAAACGGTAGAGTAATATTAATTTAAATTTAATATATATACAGCAATCAAAAGTA 18270
 QY 118 ccacagattatttattggcattgcctcactcagctaccatcagatgcgagatgat 177
 DB 18269 CCGTAATGATTATACCTTCGTCGTTTCATAAGTTCTTCGTTAAAAATTTAGGTATG 18210

QY 178 agatcatattgcatgagtcacatcttcacacaccggttggaagggt 225
 DB 18209 TAAATATTGTTTAACTAGTGAACCTAGTCTAGTGGTGGCAATGGGT 18162
 RESULT 9
 V29268
 ID V29268 standard; cDNA; 7313 BP.
 XX
 AC V29268;
 XX
 DT 21-AUG-1998 (first entry)
 XX
 DE Nucleotide sequence of human PLAG1.
 XX
 KW Human PLAG1 gene; PLAG1; tumorigenesis gene; T-gene; PLAG2; CTNNB1;
 KW antibody; benign tumour; malignant tumour; leukaemia; lymphoma;
 KW cancer; inhibition; ss.
 XX
 OS Homo sapiens.
 XX
 FH Location/Qualifiers
 CDS
 FT 481..1983
 FT /*tag= a
 FT /product= "PLAG1 protein"
 FT /*tag= b
 FT /transl_except= (pos: 1603..1605, aa: Gly)
 FT /*tag= c
 FT /transl_except= (pos: 1861..1863, aa: Gly)
 XX
 PN EP825198-A1.
 XX
 XX 25-FEB-1998.
 XX
 XX 17-JAN-1997; 97EP-0200130.
 XX
 XX 22-AUG-1996; 96EP-0202339.
 XX
 XX (KUL-) KU LEUVEN RES & DEV.
 XX (UYGO-) UNIV GOETEBORGS HOLDINGBOLAGET AB.
 XX (LEUV-) LEUVEN RES & DEV.
 PI Kas KP, Stenman KGD, Van De Ven WJM, Voz ML;
 DR WPI; 1998-132252/13.
 DR P-PSDB; W37948.
 XX
 PT New tumorigenesis T-genes and proteins - useful for, e.g. preparing
 PT antibodies for clinically diagnosing cells having non-physiological
 PT proliferative capacity such as lipoblastomas
 XX
 PS Claim 2; Fig 4; 71pp; English.
 XX
 CC This is the nucleotide sequence of the human PLAG1 gene. It is a
 CC tumorigenesis gene (T-gene), which is isolated in the form of PLAG1,
 CC PLAG2, and CTNNB1 genes. Their proteins can be used as a starting
 CC point for preparing antibodies for clinically/medically diagnosing
 CC cells having a non-physiological proliferative capacity as compared
 CC to wild type cells, where the former cells are selected from both
 CC benign and malignant tumours, as well as leukaemia and lymphomas.
 CC Derivatives of the T-gene are also used in the diagnosis and
 CC preparation of therapeutical compositions for the treatment of cancers,
 CC such as nucleic acid derivatives, and antibodies. The T-gene may be
 CC used as a starting point for designing suitable expression-modulating
 CC compounds or techniques for the treatment of non-physiological
 CC proliferation phenomena in humans or animals. Expression inhibitors
 CC of the T-gene can be used in the treatment of diseases involving
 CC benign or malignant tumours.
 XX
 SQ Sequence 7313 BP; 2287 A; 1327 C; 1356 G; 2342 T; 1 other;

RESULT 11

C skin
C syndr
C for r
C (and

CC isolating the amplified sequences. The computer readable medium can be
 CC used in a computer-based system for identifying fragments of the
 CC S. pneumoniae genome of commercial importance, or expression modulating
 CC fragments of the S. pneumoniae genome. Products from the present
 CC invention can be used in diagnosis kits and assays, and pharmaceutical
 CC compositions and vaccines for S. pneumoniae.

SQ Sequence 11887 BP; 3478 A; 2714 C; 2103 G; 3592 T; 0 other;

Query Match 12.1%; Score 28.2; DB 19; Length 11887;
 Best Local Similarity 49.7%; Pred. No. 13;
 Matches 72; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
 QY 2 ttttgcagtagtagaataatgttagtgctcctcatgataggggtggaagatttgaggtta 61
 Db 9025 TTTTAAATTCAGTATGCAATGATGGAGATCAFTATAAAACATAGAGGTCTTAAGACT 8966
 QY 62 ttgaatgaggaactctctgagtagagacttcagaaacaaatgactcgggaagctaccgc 121
 Db 8965 TTAGAATAGGCAATAAAGCATATGCAATTCCACAATATCCATGAGCAAGCTATAGTCT 8906
 QY 122 agattatttttggcattgcctc 146
 Db 8905 AAAATGCTTTCTTGGCTGTATCTC 8881

RESULT 14
 V62392
 ID V62392 standard; DNA; 12565 BP.
 AC V62392;
 XX
 DT 19-JAN-1999 (first entry)
 DE Human interleukin-1 receptor antagonist gene.
 XX
 KW Human; interleukin-1 receptor antagonist gene; IL-1; polymorphism;
 KW diagnosis; osteoporosis; ds.
 XX
 OS Homo sapiens.

XX
 PI WO9844150-A1.
 XX
 PD 08-OCT-1998.
 XX
 PF 27-MAR-1998; 98WO-GB00944.
 XX
 PR 27-MAR-1997; 97GB-0006359.
 XX

PA (GEMI-) GEMINI RES LTD.
 XX
 PI Keen RW, Spector TD;
 XX
 DR WPI; 1998-557135/47.
 XX
 PT Diagnosis of osteoporosis by determining genotype of interleukin-1
 PT receptor antagonist gene - useful for diagnosing patient
 PT pre-disposition or susceptibility to osteoporosis and for
 PT therapeutic intervention
 XX
 PS Disclosure; Page 21-27; 36pp; English.

XX A method has been developed for the diagnosis of osteoporosis comprising
 CC determining the genotype of an interleukin-1 (IL-1) receptor antagonist
 CC gene (IL-1RN). The present sequence represents the human interleukin-1
 CC receptor antagonist gene. The method can be used for the diagnosis of
 CC disease, including diagnosis of osteoporosis and predisposition or
 CC susceptibility to osteoporosis and for therapy.

SQ Sequence 12565 BP; 3217 A; 2980 C; 3072 G; 3294 T; 2 other;

Query Match 12.1%; Score 28.2; DB 19; Length 12565;
 Best Local Similarity 53.1%; Pred. No. 14;
 Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
 QY 85 tgagacttcagaaacaaatgactcgggaagctaccgcagattattttattgctgctgcc 144
 Db 12129 tgcctctgacattgtagagcttctggcatttgagacttgatgaagatgctgctgcc 12188
 QY 145 tccatcagctaccctcagctatgcagatgataagatcatattgcatgagtc 197
 Db 12189 tctgctgtctcccccagcgtggagctctgcagagcaggagaaacatgactc 12241

RESULT 15
 X75925
 ID X75925 standard; DNA; 12565 BP.
 AC X75925;
 XX

XX 29-JUL-1999 (first entry)
 DT Human interleukin 1RN gene.
 DE
 XX

XX Human; interleukin 1; IL-1B; IL-1A; IL-1RN; diagnosis; detection;
 KW chronic obstructive airway disease; chronic bronchitis; emphysema;
 KW asthma; chronic bronchiolitis; proinflammatory haplotype; ss.

XX Homo sapiens.
 XX
 PN WO9924615-A2.
 XX
 PD 20-MAY-1999.
 XX

XX 09-NOV-1998; 98WO-US23721.
 XX
 PR 12-JAN-1998; 98US-0005923.
 PR 07-NOV-1997; 97GB-0023553.
 XX

PA (MEDI-) MEDICAL SCI SYSTEMS INC.
 XX
 PI Barnes PJ, Duff GW, Giovine M, Lim S;
 XX
 DR WPI; 1999-327420/27.

XX Genotyping nucleic acid samples for interleukin-1 (IL-1)
 PT proinflammatory haplotype alleles, useful for predicting
 PT susceptibility to developing chronic obstructive airway disease
 XX
 PS Example 3; Fig 3; 37pp; English.

XX The present invention describes genotyping a nucleic acid sample from a
 CC subject to determine at least one allele of an interleukin-1 (IL-1)
 CC proinflammatory haplotype. A method has also been described for
 CC determining a subject's susceptibility to developing chronic obstructive
 CC airway disease (COAD) or for predicting the rapidity or ultimate
 CC progression of a COAD in the subject by: (a) obtaining a nucleic acid
 CC sample from the subject; and (b) detecting at least one allele of an
 CC IL-1 proinflammatory haplotype in the sample, where detection of at
 CC least one of these alleles indicates that the patient has an increased
 CC susceptibility to developing COAD. The method is useful for determining
 CC the susceptibility of subjects to developing chronic obstructive airway
 CC disease or for predicting the rapidity or ultimate progression of
 CC chronic obstructive airway disease (COAD). COAD can be asthma, emphysema,
 CC chronic bronchitis or chronic bronchiolitis. The method provides for
 CC early identification of chronic obstructive airway disease (COAD),
 CC facilitating administration of appropriate treatment at the earliest
 CC stage, thereby increasing the probability of a positive outcome. The
 CC present sequence represents the human IL-1RN gene.

SQ Sequence 12565 BP; 3217 A; 2980 C; 3072 G; 3294 T; 2 other;

Query Match 12.1%; Score 28.2; DB 20; Length 12565;

Best Local Similarity 53.1%; Pred. NO. 14;
Matches 60; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Qy 85 tgagacttcagaaacaaatgactcgggaagctaccgcagagattatttttggcattgcc 144
Db 12129 tgccttgacattgtagagcttctggcacttgagacttgatgaaagatggctgtgcc 12188
Qy 145 tccatcagtcaccatcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 197
Db 12189 tctgcctgtctccccaccaggctgggagctctgcagagcaggaaacatgactc 12241

Search completed: November 4, 2000, 13:40:19
Job time: 16146 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 4, 2000, 08:56:04 ; Search time 189.35 Seconds
(without alignments)
186.106 Million cell updates/sec

Title: US-09-300-482-1

Perfect score: 233

Sequence: 1 gttttgcagttagtagaat.....ttggacaagggtattgttg 233

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 262060 seqs, 75620727 residues

Total number of hits satisfying chosen parameters: 524120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/PCUS_COMB.seq:*

7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	29.8	12.8	7655	2	US-08-619-554-1
2	28	12.0	3417	3	US-08-464-402-1
3	27.4	11.8	5892	5	US-08-755-587-15
4	27.4	11.8	7240	5	US-08-755-587-15
5	27.4	11.8	11283	3	US-08-603-753D-3
6	27.4	11.8	11385	3	US-08-639-501-1
7	27.4	11.8	11385	5	US-09-044-946-1
c 8	26.4	11.3	1791	1	US-08-245-294-7
c 9	26.4	11.3	1791	2	US-08-474-499-7
c 10	26.4	11.3	1791	2	US-08-307-279A-7
c 11	26.4	11.3	1791	6	PCF-US95-06211-7
c 12	26.4	11.3	1917	5	US-08-755-587-1
c 13	26.2	11.2	13159	5	US-08-986-485-3
c 14	26	11.2	1294	2	US-08-464-523B-3
c 15	26	11.2	1553	5	US-08-492-459-21
c 16	26	11.2	1553	5	US-08-423-752-21
c 17	26	11.2	1776	2	US-08-464-523B-4
c 18	26	11.2	2274	5	US-08-492-459-13
c 19	26	11.2	2274	5	US-08-423-752-13
20	25.8	11.1	2733	2	US-08-676-967-3
21	25.8	11.1	2733	2	US-08-676-967-3
22	25.8	11.1	2733	4	US-09-098-487-3
c 23	25.6	11.0	1611	1	US-08-485-569-1
c 24	25.6	11.0	1611	2	US-08-480-993-1
c 25	25.6	11.0	1611	3	US-07-903-079B-1
c 26	25.6	11.0	1926	1	US-07-718-535-5

27	25.6	11.0	1926	1	US-08-161-999-5	Sequence 5, Appli
28	25.6	11.0	2106	1	US-07-718-535-4	Sequence 4, Appli
29	25.6	11.0	2106	1	US-08-161-999-4	Sequence 4, Appli
30	25.6	11.0	2106	2	US-08-161-999-4	Sequence 2, Appli
31	25.6	11.0	2277	2	US-08-676-967-2	Sequence 2, Appli
32	25.6	11.0	2277	2	US-08-676-967-2	Sequence 2, Appli
33	25.6	11.0	2277	1	US-09-098-487-2	Sequence 2, Appli
34	25.6	11.0	2657	1	US-07-718-535-2	Sequence 2, Appli
c 35	25.6	11.0	2894	1	US-08-161-999-2	Sequence 2, Appli
c 36	25.6	11.0	2894	1	US-08-278-091-1	Sequence 1, Appli
c 37	25.6	11.0	2894	1	US-08-483-859-1	Sequence 1, Appli
c 38	25.6	11.0	2894	3	US-08-472-173-1	Sequence 1, Appli
c 39	25.6	11.0	2894	4	US-08-482-816-1	Sequence 1, Appli
c 40	25.6	11.0	2894	4	US-08-296-149-1	Sequence 1, Appli
c 41	25.6	11.0	2894	4	US-08-801-499-1	Sequence 1, Appli
c 42	25.6	11.0	2894	4	US-08-615-271-1	Sequence 1, Appli
c 43	25.6	11.0	2894	5	US-09-074-660-1	Sequence 1, Appli
c 44	25.6	11.0	2894	5	US-09-074-659-1	Sequence 1, Appli
c 45	25.6	11.0	2894	5	US-09-106-468-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-619-554-1/c
; Sequence 1, Application US/08619554
; Patent No. 5821353
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS, Cameron M.
; APPLICANT: CHREBET, Gary L.
; APPLICANT: CLEMAS, Joseph
; APPLICANT: EL-SHERBEINI, Mohammed
; APPLICANT: FOOR, Forrest
; APPLICANT: KAHN, Jennifer
; APPLICANT: KELLY, Rosemarie, - PARENT, S.A.
; APPLICANT: MARRINAN, Jean, - RAMADAN, N.M.
; APPLICANT: MORIN, Nancy, - REGISTER, E.A.
; APPLICANT: ONISHI, Janet, - SHEI, Gan-Ju
; TITLE OF INVENTION: DNA ENCODING 1,3 BETA-D GLUCAN
; TITLE OF INVENTION: SYNTHASE SUBUNITS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: NEWARK
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,554
; FILING DATE: 01-AUG-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: COPPOLA, JOSEPH A
; REGISTRATION NUMBER: 38,413
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-6734
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7655 base pairs
; TYPE: nucleic acid

LENGTH: 11385 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 229..10482
US-09-044-946-1

Query Match 11.8%; Score 27.4; DB 5; Length 11385;
Best Local Similarity 65.6%; Pred. No. 8;
Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 48 aagatttgagttattgaatgaggcaatctctgagatgagacttcagaaaaaatgact 107
DB 6611 AAGAAATTAATTAATCAATTAATTAATGTTGAAGTGTCTTTCAGAAAAATAATCACT 6670
QY 108 C 108
DB 6671 C 6671

RESULT 8
US-08-245-294-7/c
; Sequence 7, Application US/08245294
; Patent No. 5644047
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DIAGNOSING
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA
; TITLE OF INVENTION: INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,294
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.612
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1649

US-08-245-294-7

Query Match 11.3%; Score 26.4; DB 1; Length 1791;
Best Local Similarity 59.2%; Pred. No. 7.9;
Matches 45; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 128 tttatttggcattgcctccatcagctcaccatcagtcgagatgataagatcatat 187
DB 1225 TTTGCTAGATCAGGACATCATTAATCTTCTCACCATTACCGAAATAATAACATCACCT 1166
QY 188 tgcattgagtcctctt 203
DB 1165 GCCTTGATACACGCTT 1150

RESULT 9
US-08-474-499-7/c
; Sequence 7, Application US/08474499
; Patent No. 5693776
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE
; TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,499
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,294
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.612
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1649
US-08-474-499-7

Query Match 11.3%; Score 26.4; DB 2; Length 1791;
Best Local Similarity 59.2%; Pred. No. 7.9;
Matches 45; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 128 tttatttggcattgcctccatcagctcaccatcagtcgagatgataagatcatat 187
DB 1225 TTTGCTAGATCAGGACATCATTAATCTTCTCACCATTACCGAAATAATAACATCACCT 1166

```

; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT-US95/06211
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,294
; FILING DATE: 18 MAY 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.6121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-9880
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1649
; PCT-US95-06211-7

Query Match 11.3%; Score 26.4; DB 6; Length 1791;
Best Local Similarity 59.2%; Pred. No. 7.9;
Matches 45; Conservative 0; Mismatches 31; Indels 0; Gaps

QY 128 tttttagtggtcctccatcgatgagtcacgcagtgcagatgagatgaagatcatat 187
||| | || | || | || | || | || | || | || | || | || | || |
DB 1225 TTGTAGTACGCGACATCAATAATCTTCACCATTACCGAAATAATACATCACCT 1166

QY 188 tgcatgagtcacctt 203
| ||| ||| |||
DB 1165 GCCTTGATACCAGCTT 1150

RESULT 12
US-08-755-587-1
; Sequence 1, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,279A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 141..1652
; US-08-307-279A-7

Query Match 11.3%; Score 26.4; DB 2; Length 1791;
Best Local Similarity 59.2%; Pred. No. 7.9;
Matches 45; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 128 tttttagtggtcctccatcgatgagtcacgcagtgcagatgagatgaagatcatat 187
||| | || | || | || | || | || | || | || | || | || | || |
DB 1225 TTGTAGTACGCGACATCAATAATCTTCACCATTACCGAAATAATACATCACCT 1166

QY 188 tgcatgagtcacctt 203
| ||| ||| |||
DB 1165 GCCTTGATACCAGCTT 1150

RESULT 11
PCT-US95-06211-7/c
; Sequence 7, Application PC/TUS9506211
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1917 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-755-587-1

Query Match 11.3%; Score 26.4; DB 5; Length 1917;
Best Local Similarity 63.9%; Pred. No. 8.1;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 48 aaggattgattgaatgagcaatctgagttgagacttcagaaacaaatgact 107
Db 176 AAGATTAAATTAATCAATACTTAATGTTGAAGTGTGTTTTCAGAAAAATACT 235

Qy 108 c 108
Db 236 C 236

RESULT 13
US-08-986-485-3/c
Sequence 3, Application US/08986485
Patent No. 6046030
GENERAL INFORMATION:
APPLICANT: WU, SHUITAN
APPLICANT: SWEET, RAYMOND
APPLICANT: TRONEH, ALEMSEGED
TITLE OF INVENTION: A HUMAN LIG-1 HOMOLOG (HLIG-1)
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/986,485
FILING DATE: 08-DEC-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/059,448
FILING DATE: 22-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F

REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70264
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3159 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-986-485-3

Query Match 11.2%; Score 26.2; DB 5; Length 3159;
Best Local Similarity 49.6%; Pred. No. 12;
Matches 67; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 83 tatgagacttcagaaacaaatgactcgggaagctaccgcagattattttttgcatg 142
Db 1776 TATGAAACCCCTGAAACCCCTTATGTACATGGATATCAATACTTTTTTTCCTTT 1717

Qy 143 cctccatcagttaccatcagttatgctgagatgataagatcatattgcatgactcat 202
Db 1716 GGATACAAATCCCTCTTGGCTTTACTGTGTTTCAGATCCCAAGTCTCTGTGAACGACT 1657

Qy 203 tcacacacccggttg 217
Db 1656 GGATACTCCACATGG 1642

RESULT 14
US-08-464-523B-3/c
Sequence 3, Application US/08464523B
Patent No. 5723761
GENERAL INFORMATION:
APPLICANT: Toni A. Voelker
APPLICANT: Ling Yuan
APPLICANT: Jean Kridl
APPLICANT: Deborah Hawkins
APPLICANT: Aubrey Jones
TITLE OF INVENTION: Plant Acyl ACP Thioesterase
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1(a)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,523B
FILING DATE: 02-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/13131
FILING DATE: 10-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/152,004
FILING DATE: 10-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/261,695
FILING DATE: 16-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Donna E. Scherer

REGISTRATION NUMBER: 34,719
NAME: Carl J. Schwedler
REGISTRATION NUMBER: 36,924
REFERENCE/DOCKET NUMBER: CGNE 100-1WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1294 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-464-523B-3

Query Match 11.2%; Score 26; DB 2; Length 1294;
Best Local Similarity 59.5%; Pred. No. 9.3;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 19 atattgtagtgcctctatgagtggaaggattgagttatgattgaatgaggaatc 78
DB 626 AAACGCTTGTGGCTGCATGACAGATTGCGCAGTTGCCAGTTGATGAACATGCCAATCAC 567
QY 79 tgagttatgagactt 92
DB 566 GGCGCATGCCATT 553

RESULT 15
US-08-492-459-21/c
Sequence 21, Application US/08492459
Patent No. 6015689
GENERAL INFORMATION:
APPLICANT: Takashi OKADO et al.
TITLE OF INVENTION: REGULATION OF AUREOBASIDIIN SENSITIVITY IN FUNGUS
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/492,459
FILING DATE: June 20, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/243,403
FILING DATE: May 16, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1553
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-492-459-21

Query Match 11.2%; Score 26; DB 5; Length 1553;
Best Local Similarity 50.8%; Pred. No. 10;
Matches 62; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 35 tatgatagggtggaaggatttgagttattggaatgaggaatctctgagttatgagacttca 94
DB 1127 TATGATAAAGGGTCAATCTCTTGGATATCGATTTTTCATTTTCAGTGTATGACCAACGA 1068
QY 95 gaaacaatgactcgggaagctaccgcagattattttatttggcatttgcctccatcagtc 154
DB 1067 CAGAAAAGGCCCTTCTTTTGGCAAAATATTTATATTGTTGTAATTCAAAAACAGTC 1008
QY 155 ta 156
DB 1007 AA 1006

Search completed: November 4, 2000, 13:33:18
Job time: 16634 sec

GenCore version 4.5
Copyright (c) 1993 - 2000, Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: November 4, 2000, 06:48:03 ; Search time 4352.3 seconds
(without alignments)
330.996 Million cell updates/sec

Title: US-09-300-482-1

Perfect score: 233

Sequence: 1 gttttgcagtagtagaat.....ttggacaagggtattgttg 233

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7189864 seqs, 3091403243 residues

Total number of hits satisfying chosen parameters: 14379728

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: em_estba.*
42: em_estfun.*
43: em_esthum1.*

44: em_esthum2.*
45: em_esthum3.*
46: em_esthum4.*
47: em_esthum5.*
48: em_esthum6.*
49: em_esthum7.*
50: em_esthum8.*
51: em_esthum9.*
52: em_esthum10.*
53: em_esthum11.*
54: em_esthum12.*
55: em_esthum13.*
56: em_esthum14.*
57: em_esthum15.*
58: em_esthum16.*
59: em_esthum17.*
60: em_esthum18.*
61: em_esthum19.*
62: em_esthum20.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estom.*
68: em_estov1.*
69: em_estov2.*
70: em_estpl1.*
71: em_estpl2.*
72: em_estpl3.*
73: em_estpl4.*
74: em_estpl5.*
75: em_estro1.*
76: em_estro2.*
77: em_estro3.*
78: em_estro4.*
79: em_estro5.*
80: em_estro6.*
81: em_estro7.*
82: em_estro8.*
83: em_estro9.*
84: em_estro10.*
85: em_estro11.*
86: em_estro12.*
87: em_estro13.*
88: gb_gss1.*
89: gb_gss2.*
90: gb_gss3.*
91: gb_gss4.*
92: em_gss1.*
93: em_gss2.*
94: em_gss3.*
95: em_gss4.*
96: gb_gss5.*
97: gb_gss6.*
98: gb_gss7.*
99: gb_gss8.*
100: gb_gss9.*
101: em_gss5.*
102: em_gss6.*
103: em_gss7.*
104: em_gss8.*
105: em_gss9.*
106: em_gss10.*
107: em_gss11.*
108: gb_gss10.*
109: gb_gss11.*
110: em_gss12.*
111: gb_gss12.*
112: gb_gss13.*
113: gb_gss14.*
114: gb_gss15.*
115: gb_gss16.*
116: gb_gss17.*

us-09-300-482-1.rst

```

Db 394 ACTGGCAAGCAATATCTGAGCAGCAATCTCAAAAAATAGCACAGAGGATCATCCAG 453
Qy 121 cagatttttttgcattgcctccatcagctacccatccatcagatgcgagatgataag 180
Db 454 AAGACTTTTCTACTTGTCTCTCGTCAGTATATCCCTCTGTTGCGAGAATGATAAA 513
Qy 181 atcatattgcattgcattccattccacacacccggttgacaaagggtttattgttg 233
Db 514 AAGTTATTGTATGAACAAATCTGATCTGTGGTGGTGGACAGCAATCGTTGTTG 566

RESULT 9
LOCUS BE449154 608 bp mRNA EST 26-JUL-2000
DEFINITION hirsutum trichome, Cornell University Lycopersicon
ACCESSION BE449154
VERSION BE449154.1 GI:9454657
KEYWORDS EST.
SOURCE Lycopersicon hirsutum.
ORGANISM Lycopersicon hirsutum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 608)
AUTHORS van der Hoeven, R.S., Bezzeredes, J.L., Matern, A.L., Holt, I.E., Liang, F.,
Fraser, C.M., Martin, G.B., Giovannoni, J.J., and Tanksley, S.D.
Generation of ESTs from wild tomato (Lycopersicon hirsutum)
trichomes
TITLE Unpublished (2000)
JOURNAL Contact: David Frisch
COMMENT Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
LOCATION/Qualifiers
Source 1..608
/organism="Lycopersicon hirsutum"
/db_xref="taxon:62890"
/clone="CLHT28B8"
/clone_lib="L. hirsutum trichome, Cornell University"
/tissue_type="trichome"
/dev_stage="mixed stages"
/note="Leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yielded a
mixture of cells which is highly enriched for trichome
likely with minor contaminations of other types of leaf
cells"
BASE COUNT 175 a 114 c 131 g 188 t
ORIGIN
Query Match 41.6%; Score 97; DB 35; Length 608;
Best Local Similarity 63.5%; Pred. No. 1.1e-19;
Matches 148; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
Qy 1 gtttttcagttagatgaatatgttagtgctcctcatgataggggtggaagatttgatt 60
Db 272 GTTTCGCAACTGATTAATACGTACGTGCTCTTATGATCTCGTGAGGGCTTTACGTT 331
Qy 61 attgaatgaggcaatctctcagatgatgagacttcagaaacaatgactcgggaagctaccg 120
Db 332 ACTGGACAAGCAATATCTGAGCAGCAATCTCAAAAAATAGCACAGAGGATCATCCAG 391
Qy 121 cagatttttttgcattgcctccatcagctacccatcagatgcgagatgataag 180
Db 392 AAGACTTTTCTACTTGTCTCTCGTCAGTATATCCCTCTGTTGCGAGAATGATAAA 451
Qy 181 atcatattgcattgcattccattccacacacccggttgacaaagggtttattgttg 233

```

```

Db 452 AAGTTATTGTATGAACAAATCTGATCTGTGGTGGACAGCAATCGTTGTTG 504
Qy 1 gtttttcagttagatgaatatgttagtgctcctcatgataggggtggaagatttgatt 60
Db 149 GTTTCGCAACTGATTAATACGTACGTGCTCTTATGATCTCGTGAGGGCTTTACGTT 208
Qy 61 attgaatgaggcaatctctcagatgatgagacttcagaaacaatgactcgggaagctaccg 120
Db 209 ACTGGACAAGCAATATCTGAGCAGCAATCTCAAAAAATAGCACAGAGGATCATCCAG 268
Qy 121 cagatttttttgcattgcctccatcagctacccatcagatgcgagatgataag 180
Db 269 AAGACTTTTCTACTTGTCTCTCGTCAGTATATCCCTCTGTTGCGAGAATGATAAA 328
Qy 181 atcatattgcattgcattccattccacacacccggttgacaaagggtttattgttg 233
Db 329 AAGTTATTGTATGAACAAATCTGATCTGTGGTGGACAGCAATCGTTGTTG 381

RESULT 11
A1941197

LOCUS A1491202 649 bp mRNA EST 29-JUN-1999
DEFINITION ESTP241911 tomato shoot, Cornell Lycopersicon esculentum cDNA clone
CLEB1G14 similar to glucose-6-phosphate dehydrogenase, mRNA
sequence.
ACCESSION A1491202
VERSION A1491202.1 GI:4386512
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 649)
AUTHORS van der Hoeven, R.S., Matern, A.L., Vision, T., Holt, I.E., Liang, F.,
Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L.,
Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Giovannoni, J.J.,
and Tanksley, S.D.
Generation of ESTs from tomato shoot meristem
Unpublished (1999)
CONTACT: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.
LOCATION/Qualifiers
Source 1..649
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEB1G14"
/clone_lib="tomato shoot, Cornell"
/tissue_type="shoot meristem"
/dev_stage="8 week old plants"
/lab_host="XLOLR"
/note="Vector: pBK CMV; Site 1: EcoRI; Site 2: XhoI; cLEB
- Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA
library made from tomato vegetative shoots including
meristems and small expanding leaves."
BASE COUNT 182 a 117 c 154 g 196 t
ORIGIN
Query Match 41.6%; Score 97; DB 11; Length 649;
Best Local Similarity 63.5%; Pred. No. 1.2e-19;
Matches 148; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
Qy 1 gtttttcagttagatgaatatgttagtgctcctcatgataggggtggaagatttgatt 60
Db 149 GTTTCGCAACTGATTAATACGTACGTGCTCTTATGATCTCGTGAGGGCTTTACGTT 208
Qy 61 attgaatgaggcaatctctcagatgatgagacttcagaaacaatgactcgggaagctaccg 120
Db 209 ACTGGACAAGCAATATCTGAGCAGCAATCTCAAAAAATAGCACAGAGGATCATCCAG 268
Qy 121 cagatttttttgcattgcctccatcagctacccatcagatgcgagatgataag 180
Db 269 AAGACTTTTCTACTTGTCTCTCGTCAGTATATCCCTCTGTTGCGAGAATGATAAA 328
Qy 181 atcatattgcattgcattccattccacacacccggttgacaaagggtttattgttg 233
Db 329 AAGTTATTGTATGAACAAATCTGATCTGTGGTGGACAGCAATCGTTGTTG 381

```

LOCUS	AT1941197	287 bp	mrna	EST	13-DEC-1999
DEFINITION	S085911.y1 Gm-c1010 Glycine max CDNA clone genome SYSTEMS CLONE ID: Gm-c1010-1821 5', similar to SW:G6PD_MEDSA_Q42919				
	GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE, CYTOPLASMIC ISOFORM ;, mRNA sequence.				
ACCESSION	AT1941197				
VERSION	AT1941197.1	GI:5698182			
KEYWORDS	EST.				
SOURCE	soybean.				
ORGANISM	Glycine max				
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.				
REFERENCE	1 (bases 1 to 287)				
AUTHORS	Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Corvelli, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.				
TITLE	Public Soybean EST Project				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Possible reversed clone: similarity on wrong strand High quality sequence stop: 106.				

BASE COUNT	73 a	43 c	73 g	96 t	2 others
ORIGIN	was constructed by Dr. Lilla Vodkin and Dr. Anu Khanna."				
Query Match	41.5%	Score 96.6;	DB 14;	Length 287;	
Best Local Similarity	63.1%;	Pred. No. 1.3e-19;			
Matches 147;	Conservative 0;	Mismatches 86;	Indels 0;	Gaps 0;	
QY	1	gttttcgagttagatgaatatgttagtgctctctatgatagggtggagagatttgatt	60		
Db	38	GTTTTACAGCTGATCAAAATATGTAGTGGCTCTTATGATTCGTGAGGATGGATTTTCGCTT	97		
QY	61	attgaatgaggcaatctcttgatgatgagactcagaaaaacaatgactcogggaagtaccg	120		
Db	98	CTTGGAATAAGAGATTTCAGAGCAATGAATTTTGAAGAAAGAGGTGTGAGGGTTTATCTCG	157		

[illegible]

BASE COUNT	216 a	130 c	146 g	205 t	
ORIGIN					
	helper phage and propagated in SOLR cells."				
	Query Match	40.5%;	Score 94.4;	DB 23;	Length 697;
	Best Local Similarity	63.8%;	Pred. No. 7.5e-19;		
	Matches 143;	Conservative 0;	Mismatches 81;	Indels 0;	Gaps 0;
QY	1	gtttttgcagtttagtagaataatttagtggctcctatgataggtgagaagatttgatt	60		
Db	474	GTTTTTACAATTGGTCAATATGTAGTGGCCCTTATGATCTGAAGATGTTTCGCTT	533		
QY	61	attgaatgaggcaatcctctgagtatgagacttcagaaaaaatgactcgggaagctaccg	120		

```
Db 534 GTTGGATAAAGAGATTTCAGAGCATGAATATTGAAAATAAGTAAAGAGGTTCACTCG 593
Qy 121 cagattatttttggttcattccatcagctcaccatcagctatgagatgataag 180
Db 594 GAGGCTTTTCTATAGCACTCTCTCTCAGTGTATCCGTCGCTTTGCAAGATGATCAA 653
Qy 181 atcatattgcattgagtcatttccacacacacccggttgacaaagg 224
Db 654 GACTGTTGGCATGAATAATCTGATCTTGTGGTGGATGACACGCG 697

RESULT 13
AW736245 489 bp mRNA EST 24-APR-2000
LOCUS EST332231 KV3 Medicago truncatula cDNA clone pkV3-1213, mRNA
DEFINITION
ACCESSION AW736245
VERSION AW736245.1 GI:7643094
KEYWORDS barrel medic.
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 489)
VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L.,
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
Fraser,C.M.
ESTs from roots of Medicago truncatula after Rhizobium inoculation
Unpublished (1999)
JOURNAL Contact: VandenBosch K
COMMENT Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 7707
Fax: 409 845 2891
Email: kate@mail.bio.tamu.edu
Texas A&M EST name: T255752e
TIGR sequence name: MPEAE50TK
More information is available at:
'http://chrysis.tamu.edu/medicago/'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES
source
1..489
/organism="Medicago truncatula"
/cultivar="genotype AL7"
/db_xref="taxon:3880"
/clone="pkV3-1213"
/clone_lib="KV3"
/tissue_type="Seedling roots"
/dev_stage="3 days post-inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain SOLR"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The
cDNA was directionally ligated into the Unizap XR vector
from Stratagene and packaged using Gigapack III Gold
packaging extracts. Plasmids containing cDNA inserts
were excised from the recombinant lambda-zap phage using
Ex-assist helper phage and propagated in SOLR cells."
BASE COUNT 132 a 94 c 113 g 150 t
ORIGIN
Query Match 40.3%; Score 93.8; DB 24; Length 489;
Best Local Similarity 62.7%; Pred. No. 1.1e-18;
Matches 146; Conservative 0; Indels 0; Gaps 0;
Qy 1 gtttttcagttagataatgttagtgcctcctatgatgaggttggaagattgagtt 60
Db 82 GTTTTACATCCGATCAATATGTAAGTGGCTCTTATGATCTGAGAAATGATTCCGTTT 141

RESULT 14
AV548140 583 bp mRNA EST 23-JUL-2000
LOCUS AV548140 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION cDNA clone RZL48allf 3', mRNA sequence.
ACCESSION AV548140
VERSION AV548140.1 GI:8719553
KEYWORDS EST.
SOURCE thale cress. thaliana
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 583)
Asanizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
JOURNAL Contact: Erika Asanizu
COMMENT The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..583
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL48allf"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 161 a 103 c 145 g 174 t
ORIGIN
Query Match 40.3%; Score 93.8; DB 19; Length 583;
Best Local Similarity 62.7%; Pred. No. 1.1e-18;
Matches 146; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 1 gtttttcagttagataatgttagtgcctcctatgatgaggttggaagattgagtt 60
Db 146 GTTTTCGACGTGATTAAATGATGAGTGGCCCTTATGATGCTGAGGAGGGTTCACAG 205
Qy 61 attgaatgaggcaatcctgtgagatgagacttcagaaacaacatgactcgggaagctaccg 120
Db 206 ATTACACAAGGCAATTTTCAGAGACGACGAAATCTCCAAAATAGTACTGAAGGGTCTTCTAG 265
Qy 121 cagattatttttggttcattccatcagctcaccatcagctatgagatgataag 180
Db 266 AAGACTGTTTATCTTTCGACTTCCACCGCTCTGTTTATCTCTCTGATGCAAGATGATCAA 325
Qy 181 atcatattgcattgagtcatttccacacacccggttgacaaagggtttattgttg 233
Db 326 GACGTGCTGCATGAACAAATCTGTATCTGGTGGATGGACACGGATTTGTTGG 378

RESULT 15
```

```
Db 534 GTTGGATAAAGAGATTTCAGAGCATGAATATTGAAAATAAGTAAAGAGGTTCACTCG 593
Qy 121 cagattatttttggttcattccatcagctcaccatcagctatgagatgataag 180
Db 594 GAGGCTTTTCTATAGCACTCTCTCTCAGTGTATCCGTCGCTTTGCAAGATGATCAA 653
Qy 181 atcatattgcattgagtcatttccacacacacccggttgacaaagg 224
Db 654 GACTGTTGGCATGAATAATCTGATCTTGTGGTGGATGACACGCG 697

RESULT 13
AW736245 489 bp mRNA EST 24-APR-2000
LOCUS EST332231 KV3 Medicago truncatula cDNA clone pkV3-1213, mRNA
DEFINITION
ACCESSION AW736245
VERSION AW736245.1 GI:7643094
KEYWORDS barrel medic.
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 489)
VandenBosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L.,
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
Fraser,C.M.
ESTs from roots of Medicago truncatula after Rhizobium inoculation
Unpublished (1999)
JOURNAL Contact: VandenBosch K
COMMENT Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 7707
Fax: 409 845 2891
Email: kate@mail.bio.tamu.edu
Texas A&M EST name: T255752e
TIGR sequence name: MPEAE50TK
More information is available at:
'http://chrysis.tamu.edu/medicago/'
Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES
source
1..489
/organism="Medicago truncatula"
/cultivar="genotype AL7"
/db_xref="taxon:3880"
/clone="pkV3-1213"
/clone_lib="KV3"
/tissue_type="Seedling roots"
/dev_stage="3 days post-inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain SOLR"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The
cDNA was directionally ligated into the Unizap XR vector
from Stratagene and packaged using Gigapack III Gold
packaging extracts. Plasmids containing cDNA inserts
were excised from the recombinant lambda-zap phage using
Ex-assist helper phage and propagated in SOLR cells."
BASE COUNT 132 a 94 c 113 g 150 t
ORIGIN
Query Match 40.3%; Score 93.8; DB 24; Length 489;
Best Local Similarity 62.7%; Pred. No. 1.1e-18;
Matches 146; Conservative 0; Indels 0; Gaps 0;
Qy 1 gtttttcagttagataatgttagtgcctcctatgatgaggttggaagattgagtt 60
Db 82 GTTTTACATCCGATCAATATGTAAGTGGCTCTTATGATCTGAGAAATGATTCCGTTT 141

RESULT 14
AV548140 583 bp mRNA EST 23-JUL-2000
LOCUS AV548140 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
DEFINITION cDNA clone RZL48allf 3', mRNA sequence.
ACCESSION AV548140
VERSION AV548140.1 GI:8719553
KEYWORDS EST.
SOURCE thale cress. thaliana
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 583)
Asanizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
JOURNAL Contact: Erika Asanizu
COMMENT The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asanizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..583
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZL48allf"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/notes="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 161 a 103 c 145 g 174 t
ORIGIN
Query Match 40.3%; Score 93.8; DB 19; Length 583;
Best Local Similarity 62.7%; Pred. No. 1.1e-18;
Matches 146; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 1 gtttttcagttagataatgttagtgcctcctatgatgaggttggaagattgagtt 60
Db 146 GTTTTCGACGTGATTAAATGATGAGTGGCCCTTATGATGCTGAGGAGGGTTCACAG 205
Qy 61 attgaatgaggcaatcctgtgagatgagacttcagaaacaacatgactcgggaagctaccg 120
Db 206 ATTACACAAGGCAATTTTCAGAGACGACGAAATCTCCAAAATAGTACTGAAGGGTCTTCTAG 265
Qy 121 cagattatttttggttcattccatcagctcaccatcagctatgagatgataag 180
Db 266 AAGACTGTTTATCTTTCGACTTCCACCGCTCTGTTTATCTCTCTGATGCAAGATGATCAA 325
Qy 181 atcatattgcattgagtcatttccacacacccggttgacaaagggtttattgttg 233
Db 326 GACGTGCTGCATGAACAAATCTGTATCTGGTGGATGGACACGGATTTGTTGG 378

RESULT 15
```


AV561737
 LOCUS AV561737 624 bp mRNA EST 23-JUL-2000
 DEFINITION AV561737 Arabidopsis thaliana green siliques Columbia Arabidopsis
 thaliana cDNA clone SQ156f07F 3', mRNA sequence.
 ACCESSION AV561737
 VERSION AV561737.1 GI:8733163
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 1 to 624)
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp; URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1..624
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ156f07F"
 /clone_lib="Arabidopsis thaliana green siliques Columbia"
 /tissue_type="green siliques"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 168 a 108 c 153 g 195 t
 ORIGIN

Query Match 40.3%; Score 93.8; DB 19; Length 624;
 Best Local Similarity 62.7%; Pred. No. 1.1e-18;
 Matches 146; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 1 gtttttcagttagtagatatttagtggtcctctatagaggggtggaagatttgatt 60
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 207 GTTCTGCAGCTGATTAGTATGTGAGTGGCCCTTATGATGCTGAGGAGGGGTTCCAGAG 266

QY 61 attgaatgaggcaatctctgagatgagacttcagaaacaaatgactcgggaagctaccg 120
 ||| | |||| || || || || || || || || || || || || || || || || ||
 Db 267 ATTAGACAGGCAATTTCAGACGACGAAATCTCCAAAATAGTACTGAAGGGTCTTCTAG 326

QY 121 cagattattttatttgccattgcctccatcagtcacccatcagtcagtcagagataaag 180
 ||| | |||| || || || || || || || || || || || || || || || || ||
 Db 327 AAGACTGTTTATCTTGCACTTCCACCGCTCTTTATCTCTCTGATGCAAGATGATCAA 386

QY 181 atcatatgcatgagtccttcacacacccggttgacacaggggttatgttg 233
 || | |||| || || || || || || || || || || || || || || || || ||
 Db 387 GACGTGCTGCATGACAAATCTGATCTTGGTGGATGGACACGGATTGTTGTGG 439

Search completed: November 4, 2000, 11:51:59
 Job time: 18236 sec

